The Dissertation Committee for Luke Adam Decker
certifies that this is the approved version of the following dissertation:

Parameter selection in seismic data analysis problems

Committee:

Sergey B. Fomel, Supervisor

Todd Arbogast

Omar Ghattas

Douglas Foster

Mary Wheeler

Parameter selection in seismic data analysis problems

by

Luke Adam Decker

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2021

For Elvira

Acknowledgments

To everything there is a season,

and a time for every purpose under heaven

— Ecclesiastes 3:1

Life is full of unforeseen twists and turns. When I graduated from Whitman College with a Physics degree as a 22 year old at the start of the Great Recession, my biggest desire was to find a job working in a winery in Walla Walla, the then sleepy college and prison town now turned wine destination where I had studied. None of them were hiring. Thankfully, my roommate and great friend John Hook had recently started working as a mudlogger, or wellsite geotechnician on drilling rigs, for a small geotechnology service company and buffalo ranch called King Canyon Buffalo. Back then, oil was at $140 a barrel, and there were entry level opportunities. John was able to get me a job as a sample catcher, which began my career in geosciences. With his help and the help of the owner, Chris Nerud, wellsite geologist Phil Littlefield, and geotechs Kevin Ellis and Ryan Kranz I was able to learn the ropes and advance in the company. Every day was different. The work was interesting, the money was good, and the people were great. However, I realized that long term oilfield life wasn't for

me, so I decided to return to school at Western Washington University.

I am grateful for the year I spent at Western earning a degree in geophysics. My adviser, Jackie Caplan-Auerbach was incredible, giving me time and support as if I was one of her graduate students and organizing an exciting geophysical research cruise off the coast of Hawaii's Big Island where I discovered I am a landlubber. Other faculty members like Bernard Housen, Scott Linneman, and Liz Schermer took great joy in teaching and instilled a passion for earth science. I have been very lucky with undergraduate professors, at Whitman my adviser, Kurt Hoffman, and others like Dana Burgess, Doug Juers, Albert and Laura Schueller, and Fred Moore were similarly excellent. And I can not forget Pat Langhorne, my adviser for a year at The University of Otago's Sea Ice Research Group, who provided me with my first research experience.

After Western I came to Austin, beginning seven non-continuous years in my adviser Sergey Fomel's Texas Consortium for Computational Seismology at The University of Texas, first as a Master's student in the Jackson School Geosciences, and later as a PhD student in the Oden Institute for Computational Engineering and Sciences. I cannot overstate my gratitude for Professor Fomel, I owe much of what I have accomplished to his patient, brilliant guidance and support. As a member of TCCS I have had the distinct privilege of befriending, learning from, and working with some incredibly brilliant people, including Ray Abma, Salah Al-Hadab, Sean Bader, Lubna Barghouty, William Burnett, Yangkang Chen, Hanming Chen, Jiubing Cheng, Hector Corzo-Pola, Gang Fang, Medhi Far, Shuang Gao, Zhicheng Geng, Shaunak Ghosh, Xufei Gong, Sarah Greer, Ben Gremillion, Jingwei Hu, Kristian Jensen, Jun

Ji, Parvaneh Karimi, Harpreet Kaur, Siwei Li, Nam Pham, Mason Philips, Kelly Regimbal, Karl Schleicher, Yunzhi Shi, Xiaolei Song, Yanadet Sripanich, Yuhan Sui, Junzhe Sun, Ryan Swindeman, Xiaokai Wang, Xinming Wu, Zhiguang Xue, Yunan Yang, and Tieyuan Zhu. I am particularly indebted to my frequent coauthors and collaborators, Alexander Klokov and Dmitrii Merzlikin, and my mentor both as a student and intern, Vladimir Bashkardin.

As a student at The University of Texas I have had so many wonderful professors. First I would like to thank my dissertation committee, Professors Sergey Fomel, Todd Arbogast, Omar Ghattas, and Mary Wheeler, and Dr. Douglas Foster for evaluating and improving my research and providing essential support and guidance along the way. I am also grateful to professors George Biros, Tan Bui, Clinton Dawson, Leszek Demkowicz, Dmitrii Makarov, Robert Moser, and Kui Ren for instructing wonderful classes, and to Nathan Bangs, Charlie Dey, Tom Hess, Xavier Janson, Lars Koesterke, Mrinal Sen, Kyle Spikes, and Clark Wilson for inspiring conversations. The help of Melissa Coffman, Phillip Guerrero, Jennifer Edwards, Stephanie Rodriguez and Jessica Rowling was invaluable for navigating the administration of graduate school life.

Over the last eight years I have been blessed to have the opportunity to work as an employee or intern at Anadarko Petroleum Corporation, Chevron, Repsol, and Numerical Algorithms Group, and had incredible mentors and made great friends. I am particularly indebted to Peeter Akerberg, who tragically passed this year. He was one of the few people I have met whose brilliance, patience, humor, and compassion can match my adviser's. I am also grateful for the industrial mentorship

of Scott Baker, Nick Battaglino, Errol Blumenthal, Sarah Cooke, Joe Dellinger, Loralee Dickson, Jim DiSienna, Kathy Dull, John Grant, Keith Gray, Ross Hill, Richard Hook, Sam Kaplan, Qingbo Liao, German Larrazábal, Rob Meyer, Gary Murphy, Jeff Nealon, Francisco Ortigosa, Adrian Peinado, Fred Shirley, Geoff Vogel, John Washborne, Anar Ysifov, and Qunshan Zhang. I hope all my friends and colleagues from the Seismic Acquisition Services Team are doing well wherever they are, and thank them for the wonderful, though brief, time we shared.

Being a member of the Society of Exploration Geophysicists and European Association of Geoscientists and Engineers has provided me with an avenue to meet numerous enriching mentors and friends. I have been lucky to know all the members of the Emerging Professionals International Committee, The Health, Safety, Security, and Environment Committee, and the Distinguished Lecturer Committee, particularly Gustavo Carstens, Paul Cunningham, Rocco Detomo, Kathryn Elkins, Joe Estep, Francesca Fazzari, Huw James, Michael Lis, John Meade, Robert Merrill, Brandy Michael, Aurelian Roeser, Gerald Schuster, Jeff Shragge, Ian Threadgold, and Johnna Yoder. And I will never forget the week in St. Petersburg with Tariq Allkhalifah, Gladys Gonzales, and Paul Sava shortly after the annexation of The Crimea, instructing a `Madagascar` software class during the days and afterwards touring Peterhof and The Hermitage, feasting on Georgian food, and enjoying the local jazz clubs as some of the few foreigners in the city at the time.

I am grateful to my friends for the strength and support they have provided me. Thank you Mariajo Acevedo, Finn Åkr, Mario Alcolea, Angel Ruiz Alonso, Lauren Ames, Iñigo Arregui, Ekain Arrizabalaga, Dan Arnost, Fernando Avila, Curt

Baldwin, Elena Barbero, Bill Beaupre, Lauren Becker, Peter Carlson, Mickey Cassar, James Cherrie, Jhonny Cisneros, Sarah Coyle, Ivana Escobar, Tatiana Gallego, Alejandro Garcia, Nick Garneski, Gabriel Gavidia, Vinicius Gayer, Julian Gerhart, Lucy Gilster, Chris Ginn, Ana Girones, Eric Goldfarb, Eva Maria Gomez, Rachel Goodman, Heather Grant, Patrick Gustie, Jack Ferrel, Todd Hawes, Stefan Henneking, Scott Hunthausen, Eric and Hannah Hutton, Nat Jacob, David Ju, Cassie Kamischke, Altina and Kazbek Kazhyen, Michael Kelly, Jack and Courtney Kenning, Tyler Kent, Jon Kerr, Kristin and Tyler Klansnic, Tredic Leichner, Ryan Lewis, Curt Lindley, Joyce Liy, Joel Lunsford, Ed Marshall, Alex Martín, Sam Martinez, John Mateer, Wes Matlock, Eric McAlvey, Dara Merz, Joe Miller, Sam Morris, Karl Nacalaban, Danielle Neihls, Liz Oberhausen, Michael O'Connor, Jovier Oliva, David Pantoja, Michael Patterson, Max and Kelsey Pedersen, Ian Pilgrim, Marcos Porteiro, Jason Ramos, Paul Reeves, Jake Rogers, Patricia Rubis, John Salvatier, Sean Scott, Sheroze Sheriffdeen, Jessica Smith, Fin Sparks, Keith and Rachel Strand, Alec Sugar, Jamey Stutz, Allison Solanki, Ariel Spenser, John Swartz, Edward Thenell, Eugen Tudor, Kento Ushikubo, Damian Whitburn, Alex and Tommy Williams, Charlie Wyman, and Polyanna and Ryuhei Yokokawa. Thank you to all the old boys and girls of the Whitman College and Otago Rugby Football Clubs, and the many wonderful people I have met and shared time with bicycle touring.

I am blessed to have a wonderful family. Thank you to my wife, Elvira, for your patience, understanding, and support during the last five years. You are the most important person in my life. Thank you to my parents, Gary and Betty, and my brother, Adam. Thank you Aunt Maureen, Aunt Pam, Aunt Sue, Uncle

Dalen, Uncle Joe, Uncle Tony, my cousins Carter, Josh, Kellen, and Galen, and my wonderful welcoming in-laws Christina, Hector, Mariano, and Pilar. Thank you also to my godfather Fred. And though they are no longer here I send thanks to my grandparents Donald, Joseph, Margaret, and Patricia.

Thank you so much to everyone who has made this dissertation possible. Knowing each of you has enriched my life and made me a better person.

Luke Decker

The University of Texas at Austin

May 2021

Parameter selection in seismic data analysis problems

Luke Adam Decker, Ph.D.
The University of Texas at Austin, 2021

Supervisor: Sergey B. Fomel

Seismic imaging is an essential tool for non-invasive subsurface evaluation. It enables Earth scientists to create a picture of the planet's interior, predicting the rocks and structures that lie below. This can enable characterization of tectonic margins to better understand the deep history of the planet, delineation of aquifers to provide water, and the safe and economic exploration for commercial oil and gas accumulations for energy production.

To generate these images numerous observations of the subsurface are taken and they are transformed to a common domain where observations of the same point in the subsurface overlay. These transformations typically are linear on the observed data and usually depend on a parameter related to seismic wave propagation, like the speed at which a seismic wave travels through the subsurface, in a non-linear manner. Selecting and determining these parameters is a crucial step in the generation of seismic images. Using inaccurate parameters in the transformations involved in seismic data processing results in seismic images that are distorted, inaccurate representations

of the subsurface. Because these parameters are related to seismic wave propagation, their values can provide insight into the composition of the Earth's interior, including the rocks or fluids present.

In this dissertation, I present methods for accurately determining those parameters and how they may be used to efficiently generate accurate, well resolved images of the Earth's interior. I show how dynamic time warping may be used to create an operator which efficiently corrects for the blurring and distortion present in seismic images caused by seismic anisotropy, or wave propagation speed changing with the direction of travel, while simultaneously characterizing and quantifying that anisotropy. I demonstrate how slope-decomposed seismic images may be transported along their characteristics in a process called oriented velocity continuation to efficiently generate a suite of images over a range of plausible migration velocities, and how oriented velocity continuation may be used with seismic diffraction imaging to determine migration velocity. The use of oriented velocity continuation is further expanded on to generate a framework for probabilistic diffraction imaging using a collection of weights computed from slope-decomposed images that represent the probability of a correctly imaged diffraction existing at a point in space for a given migration velocity, while simultaneously outputting the most likely migration velocity at each point in space. This method generates seismic images with significantly improved signal to noise ratios compared to conventional approaches. Finally, I formulate a variational method for picking an optimal surface representing how a parameter evolves in space from a volume representing the quality of fit for different parameter values based on iteratively minimizing a functional. I prove that minimiz-

ers for that functional exist, and that an iterative method will converge to a minimizer in an infinite dimensional setting. The method is applied using continuation, or graduated optimization, to avoid local minima and used to determine seismic velocities as a component of seismic processing workflows and perform automatic interpretation of a seismic horizon.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

MOTIVATION

The last century has witnessed incredible growth for the world economy and an accompanying increase in energy demand which has primarily been met by fossil fuel sources, particularly oil and gas (Roser, 2013; Feenstra et al., 2015; BP, 2020). Energy demand is forecast to continue to increase. Although non-hydrocarbon based sources are providing an increasing share of the global energy market, hydrocarbon based energy still dominates production, and demand is expected to continue to rise (OPEC, 2020).

To meet this growing demand for energy, the oil and gas industry has continued exploring for commercially viable hydrocarbon deposits, and devoted large amounts of resources to locating, delineating, and extracting these deposits in a safe manner. Accurately evaluating these subsurface prospects and assets with minimal capital expenditure constitues a key component of this process. Geophysical methods, particularly seismic imaging, provide an invaluable tool for petroleum exploration, as well as key component of optimizing efficient production of a developed petroleum deposit (Newendorp, 1976; Sheriff and Geldart, 1995; Brown, 2011).

TECHNICAL BACKGROUND AND CONTEXT

Seismic imaging transforms the echoes of waves passing through the earth into a representation of subsurface structure (Claerbout, 1971). Collecting the seismic data used in this the process typically involves setting off an explosive, vibrating a metal plate, or firing a high pressure air cannon in water, known as a "shot", and then setting out an array of sensors or "receivers", either planted geophones on land or a towed line of hydrophones at sea, and using magnetic induction in the receiver to tranform the vibrations recorded by those sensors after the each shot to an electric potential. This potential is converted to a digital representation and recorded (Vermeer, 2012).

Recorded seismic data are typically noisy. To reduce noise, many shot-receiver pairs are used with different distances or angles between those pairs, and their records are transformed to some common domain where signals representing a reflection at the same point in space are aligned. Thus their recorded waveform representing a reflection will interfere constructively and appear when the signals are averaged or "stacked" (Claerbout, 2008). One common example of this is the normal moveout with offset correction (NMO) and the dip moveout with offset correction (DMO), which transform recorded seismic data featuring following different ray paths based on subsurface velocity or reflector slope so that the records of reflection events are aligned (Hale, 1984; Yilmaz, 2001). Another is seismic migration, which transforms events recorded seismic data to the position where they occurred, either in the time or depth domain, creating an image of the Earth's interior (Gazdag, 1978; Stolt, 1978; Claerbout, 1985; Hill, 1990; Yilmaz, 2001). Seismic imaging can also be used

to create a representation of the small scale geologically interesting features within the Earth's interior like faults, fractures, pinch outs, karst, and voids through the process of seismic diffraction imaging, which provides another definition for migration – correctly applied seismic migration maps diffraction energy to its scattering point (Harlan et al., 1984; Fomel et al., 2007; Moser and Howard, 2008; Klokov and Fomel, 2012; Decker et al., 2015).

To illustrate a simple example of how these corrections operate, I now examine the NMO correction. This correction approximates the travel time of the reflection off a flat surface centered halfway between a shot and receiver

$$t \approx \sqrt{t_0^2 + 4\frac{h^2}{v^2\left(t_0\right)}}, \tag{1.1}$$

where t is the observed travel time, v is the moveout velocity of the wave through the subsurface, and half offset, h is half the distance between the source and receiver (Yilmaz, 2001). Zero-offset travel time, t_0, is the amount of time a seismic wave would take to travel from the source, to the reflector, to the receiver if the source and receiver were at the same position immediately above the reflector. For depth z and velocity $v(z)$, the zero-offset travel time t_0 may be expressed as:

$$t_0 = 2\int_0^z \frac{d\xi}{v\left(\xi\right)}. \tag{1.2}$$

If a seismic processor has a collection of seismic records who share the same midpoint between shot and receiver, called a Common Midpoint (CMP) gather, transforming each shot record from the observed t to t_0 using the appropriate offset, x and

NMO velocity v should create a collection of records whose reflections are aligned and add constructively when an average of the signals is computed.

NMO correction is an example of a parameterized operation in seismic processing. Using a parameter, NMO velocity v, it is able to transform data to a domain where features are aligned, signals are enhanced, and thus subsurface characterization is aided. In order for the method to function properly the parameter it utilizes must be accurate. Numerous essential operations in seismic processing, including the DMO correction and seismic time and depth migration mentioned above function similarly – they act as a linear transformation for seismic data that depend in some (often non-linear way) on one or more parameters which typically vary in space (Yilmaz, 2001).

Determining accurate values to use in these parameterized operations is an essential step in seismic processing. Inaccurate values may severely distort the representations of the Earth's interior they create, leading geoscientists to believe that subsurface structures that could trap hydrocarbons exist when none do, or completely miss an existing and commercial accumulation. Seismic processing performed with inaccurate values could also indicate that an area is free from geological hazards to drilling or operations when such features exist, hindering successful risk and hazard mitigation. The parameters themselves can be useful for subsurface characterization. For example rock type, the presence of hydrocarbon, and even the porosity of a reservoir may be determined from seismic velocity information (Mavko et al., 2020). The parameterized operations may be computationally expensive to apply, and seismic datasets can be quite large, often exceeding terrabyte size. Thus, seismic

processing practitioners are motivated to seek methods that accurately approximate the effects of more expensive operations with reduced computational cost. A popular example of this is seismic time migration, which uses approximate Green's functions and approximations for seismic ray travel time to provide an estimate for the outputs of the more expensive seismic depth migration (Zhang and Zhang, 1998; Fomel and Kaur, 2021). Similarly, different flavors of seismic depth migration exist depending on the computational resources a seismic processer is willing to dedicate to modeling wave propagation. Seismic waves traveling through the subsurface often propagate at different speeds depending on their orientation. This effect is intrinsic to the material the wave is propagating through, is referred to as seismic anisotropy, and may be modeled by one or more parameters depending on the situation (Crampin, 1981; Thomsen, 1988; Grechka et al., 2005; Tsvankin et al., 2010). Fully accounting for anisotropy in seismic processing can result in significant additional computational expense. Failing to account for it may lead to distorted or lower resolution images in study areas where anisotropic effects are pronounced.

TECHNICAL CONTRIBUTIONS

This dissertation focuses on developing techniques for determining parameters used in seismic processing workflows, and accurate approximations for more computationally expensive operations. These techniques used to create algorithms which are applied to field datasets. Field data applications illustrate how the determined parameters can be used to both characterize the subsurface through the physical properties they represent, like seismic velocity, and generate enhanced representations of the Earth's

interior that better resolve subsurface structures and small scale geologically interesting features.

Below I summarize the technical contributions of this work.

1. Developing an algorithm based on dynamic programming to efficiently correct for the effects of anisotropy in seismic depth images, determine the fastest axis of wave propagation, and provide a measure of the relative difference in velocity between the fastest and slowest wave propagation orientation.

2. Presenting a method for efficiently creating a collection of seismic images over a range of seismic velocities by transporting slope decomposed images along their characteristics, and illustrating how this method may be used in conjunction with seismic diffraction imaging to determine migration velocity with limited offset data.

3. Showing how a collection of slope decomposed seismic diffraction images created using a range of velocities may be used in conjunction with path-integral imaging to highlight features with a high likelihood of being seismic diffractions, suppress noise and other signal which is not likely related to diffraction, and automatically output the most likely seismic velocity as well as a measure of confidence in that velocity.

4. Proposing a variational method which is able to determine the best fit parameter surface from a volume measuring parameter fit quality.

THESIS OUTLINE

This dissertation is organized in the following manner:

- **Chapter 1: Introduction**

  Chapter 1 identifies the motivation and research objectives of this work. Context is provided for the problem, and the technical contributions are summarized. This chapter also holds the dissertation's outline.

- **Chapter 2: Quantifying and correcting residual azimuthal anisotropic moveout in image gathers using dynamic time warping**

  Chapter 2 shows how the dynamic programing technique of dynamic time warping may be used to inexpensively correct unaccounted for anisotropy in seismic depth images. This chapter also provides a mathematical treatment of how the action of dynamic time warping on data can be used to predict the dominant anisotropic axis and a measure of the strength of anisotropy present in an area within a seismic volume. An algorithm to accomplish this task is implemented in a massively parallel high performance setting using MPI so it may be efficiently used large datasets and applied to a field dataset. Use of the algoritm produces higher bandwidth seismic images with more coherent seismic reflections and better focused energy, and creates predictions of anisotropy that mesh with observations from drilled wells. I am grateful to Equinor, Repsol, and Geophysical Pursuit, Inc for permission to use the field data presented in this chapter.

- **Chapter 3: Diffraction imaging and time-migration velocity analysis using oriented velocity continuation**

  Chapter 3 introduces oriented velocity continuation, a method for efficiently generating a suite of slope-decomposed seismic images over a range of migration velocities given an initial image. A mathematical treatment of oriented velocity continuation is presented, and an algorithm to perform the method is efficiently implemented in the Fourier domain. This chapter additionally shows how oriented velocity continuation may be applied to seismic diffraction imaging to determine migration velocities by selecting the velocity that maximizes the coherence of migrated slope-decomposed seismic diffraction events. Applying the approach to synthetic and field data shows its utility for seismic processing.

- **Chapter 4: A probabilistic approach to seismic diffraction imaging**

  Chapter 4 expands on the concept of oriented velocity continuation introduced in Chapter 3 by applying the concept of path-integral imaging. Path-integral imaging uses a linear combination of images with generated with different velocities at each imaged point to create a seismic image. Treating this image as an expectation value inspires the creation of path-integral weights tied to the likelihood of correctly migrated diffraction occurring at a location within a seismic image given a migration velocity. Using these weights enables both the creation of seismic diffraction images which significantly suppress noise and non-diffractive features while amplifying diffraction, and the generation of the most likely seismic migration velocity at each point in space along with a measure of certainty in that velocity. A toy model example illustrates the concepts

this imaging method employs, applying the method to a synthetic dataset shows how it is able to significantly reduce noise in seismic diffraction images relative to other methods, and finally using the approach as part of a seismic processing workflow on a field dataset demonstrates the utility of the method in practice.

- **Chapter 5: A variational approach for picking optimal surfaces from semblance-like panels using continuation**

  Chapter 5 proposes a variational method for extending one dimensional methods for selecting a dominant trend through a panel to picking a dominant surface through a volume by minimizing a functional. This formulation ensures that spatially adjacent information is incorporated when finding that surface, and generates continuous surfaces which are continuous without directly imposing smoothing on the surface. A mathematical treatment of the method is presented, and it is proven that a minimizer exists, and that an iterative method will converge to a minimizing surface in an infinite dimensional setting given some assumptions on the behavior of critical points for the functional. A high performance computer program for computing the gradient of the picking functional is developed using Numba, and used with a two loop recursion limited-memory Broyden–Fletcher–Goldfarb–Shanno (ℓ-BFGS) algorithm to accelerate convergence. Adopting a continuation, or graduated optimization, approach enables the method to avoid many local minima and find a surface with a lower associated cost. The method is used to perform velocity analysis on two field datasets as part of seismic processing workflows, and the ability of continuation to reduce the dependance of the method on starting model is demonstrated.

To illustrate the versatility of the approach beyond selecting best fit seismic parameter surfaces it is used to perform automatic interpretation of a horizon in a 3D seismic image.

- **Chapter 6: Conclusion**

  Chapter 6 concludes this dissertation by summarizing the research presented its technical contributions. Promising directions for future extension of the work are identified.

Chapter 2

Quantifying and correcting residual azimuthal anisotropic moveout in image gathers using dynamic time warping

I propose and demonstrate a novel application of dynamic time warping (DTW) for correcting residual moveout in image gathers, enhancing seismic images, and determining azimuthal anisotropic orientation and relative intensity when moveout is caused by wave propagation through a medium possessing elliptical horizontally transverse isotropy (HTI). The method functions by first using DTW to determine the sequences of integer shifts that most closely match seismic traces within an image gather to the its stack, and then applying those shifts to flatten the gather. Flattening shifts are fitted to an ellipse to provide an approximation for the orientation and

[1]Some of the material in this chapter was published as:

- Decker, L., and Q. Zhang, 2019, Correcting residual HTI moveout and determining principal anisotropic azimuth in arbitrarily sampled image gathers using dynamic time warping: SEG Technical Program Expanded Abstracts 2019, 404–408. The authors contributed to this paper as follows: study conception and design: L. Decker and Q. Zhang; data collection: L. Decker; analysis and interpretation of results: L. Decker and Q. Zhang; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.
- Decker, L., and Q. Zhang, 2020, Quantifying and correcting residual azimuthal anisotropic moveout in image gathers using dynamic time warping: Geophysics, **85**, O71–O82. The authors contributed to this paper as follows: study conception and design: L. Decker and Q. Zhang; data collection: L. Decker; analysis and interpretation of results: L. Decker and Q. Zhang; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.

relative strength of elliptical HTI anisotropy. I demonstrate the method on synthetic and 3D field data examples to show how it is able to (1) correct for residual azimuthal anisotropic moveout, (2) accurately recover high-frequency information and improve feature resolution in seismic images, and (3) determine the anisotropic orientation while providing a measure of relative strength of elliptic anisotropy. While the method is not intended to replace anisotropic processing techniques for moveout correction, I find that it has the ability to inexpensively approximate the effects of such operations while providing a representation of the elliptic HTI anisotropy present within a volume.

INTRODUCTION

The upper crust of the Earth is a complex and heterogeneous media containing different rock strata with constituent bedding planes, fracture networks, faults, and other features below seismic resolution that may cause the measurement of elastic properties to change with orientation. In such media, seismic waves propagate at different velocities depending on their direction of travel (Crampin, 1981, 1984a, 1985; Thomsen, 1988). This phenomena is referred to as seismic anisotropy, and there is a rich tradition of geophysical literature and research focused on its modeling and processing (Crampin, 1984b; Helbig, 1994; Thomsen, 2002; Helbig and Thomsen, 2005; Grechka, 2009; Tsvankin et al., 2010; Tsvankin, 2012).

Anisotropy type is determined by the symmetries present for wave propagation as a function of orientation. Transverse isotropy (TI) models media possessing a single axis of rotational symmetry (Crampin, 1986; Thomsen, 1988). Vertically transverse

isotropy (VTI) refers to situations where horizontal wave propagation has the same velocity regardless of azimuthal orientation. In these circumstances velocity varies with inclination relative to bedding. This case is found to effectively model shales. Allowing the symmetry axis to tilt with dipping beds gives rise to tilted transverse isotropy (TTI). Horizontally transverse isotropy (HTI) occurs when the symmetry axis tilts fully to the horizontal. This type of anisotropy effectively models rocks with vertically aligned fracture networks and may be used to predict the fracture network orientation (Corrigan et al., 1996; Tod et al., 2007). Situations where velocity depends on both the inclination and azimuth of propagation lead to orthorhombic anisotropy and lower symmetry anisotropy systems. In the case of orthorhombic anisotropy the medium possesses three planes of symmetry rather than symmetry axes (Crampin, 1986; Thomsen, 1988; Tsvankin, 1997).

Orthorhombic anisotropy may be completely described by nine parameters, inverting for all of which becomes a computationally expensive exercise. This motivates the approximation of wave propagation with simpler anisotropy models possessing fewer parameters (Grechka et al., 2005), such as TI models which may be fully described with five (Tsvankin, 1997, 2012). Stronger assumptions, namely that only vertical and horizontal velocities differ, allow approximation of P-wave anisotropy with just a single anisotropy parameter (Alkhalifah and Tsvankin, 1995). An approximation for HTI media wave propagation provided by Grechka and Tsvankin (1998) shows that the variation in velocity as a function of azimuth is elliptical. This type of HTI anisotropy is known as *elliptical HTI anisotropy*, and occurs when the phase slowness and group velocity surfaces are ellipsoidal. The kinematics of elliptical HTI

anisotropy may be described by two anisotropy parameters plus velocity for each subsurface position (Abedi et al., 2019).

Seismic processing techniques incorporating anisotropy, which thus involve inverting for anisotropic parameters, are more computationally expensive than those that do not (Alkhalifah et al., 1996). Techniques that allow for more complex anisotropy, featuring additional parameters and thus additional degrees of freedom, are still more expensive. Nonetheless, failing to fully account for anisotropy leads to seismic images that are less focused and accurate (Helbig and Thomsen, 2005). This is because many traces are migrated with what amounts to an incorrect velocity (Alkhalifah and Larner, 1994; Thomsen, 2001). The sensitivity of depth domain imaging techniques to velocity perturbation further accentuates these effects in depth images (Tsvankin et al., 2010). Seismic processing workflows must therefore balance the demands of accuracy and efficiency when treating anisotropy, and seismic processing practitioners are motivated to seek approximations that are able to account for anisotropy in a more efficient way (Helbig and Thomsen, 2005; Tsvankin et al., 2010). This desire prompted Burnett and Fomel (2009) to formulate a velocity - independent method for correcting azimuthal velocity variations using local traveltime slopes in common midpoint (CMP) gathers. I propose to use dynamic time warping to inexpensively perform the correction.

The digital signal processing technique of dynamic time warping (DTW), developed by Sakoe and Chiba (1978) and applied to seismic imaging problems by Hale (2013), determines the set of integer shifts, $s[i]$, for a signal sample index i, that most closely align a matching signal, $g[i]$, to a reference signal, $f[i]$, such that

$f[i] \approx g[i + s[i]]$. Constraints are applied to the process by declaring the maximum possible shift and the maximum strain, or how quickly shifts are permitted to change with respect to index i. Shifts $s[i]$ are determined by selecting a set of integer shifts obeying the imposed constraints that minimize the accumulated mismatch between the reference and matching signals over their entirety. Because only integer shifts are considered, calculation is a relatively rapid process.

When used in conjunction with a seismic migration method accounting for a VTI media, DTW enables me to correct for the residual azimuthal anisotropic moveout resulting from the un-accounted for elliptical HTI anisotropy in a computationally efficient manner. The algorithm is unconcerned with the physics of wave propagation and treats this moveout correction as a less expensive integer-shift data matching problem. If I further assume that the HTI fast-axis is aligned parallel to fracture networks present in the subsurface, determining the principal axes of the anisotropic ellipse with respect to wave propagation azimuth will provide a sort of average fracture network orientation over the whole ray path. Similarly, the elongation of this ellipse provides a notion of how anisotropic the material is over the whole ray path.

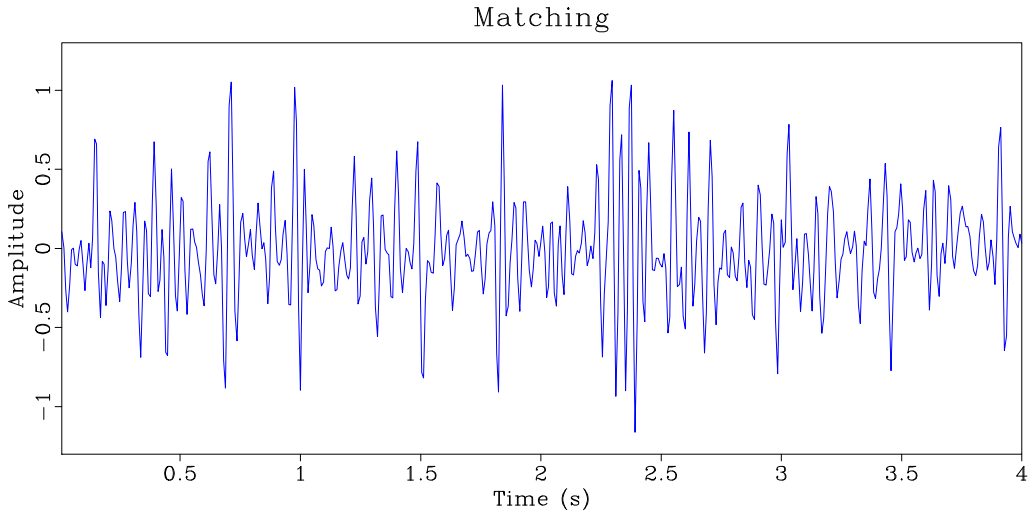
In the following sections this paper details how DTW may be used to create a method that compensates for residual elliptic HTI anisotropy in image gathers, resulting in enhanced seismic images and providing a measure of the orientation and relative intensity of that anisotropy, as well as the assumptions made in developing the method. The ability of the proposed technique to successfully recover high-frequency signal and the principal anisotropic axis orientation is demonstrated on a synthetic gather featuring residual moveout caused by artificial elliptical HTI anisotropy. The

method is then applied to a 3D field data set, generating sharper, more coherent images, as well as plausible information about HTI anisotropy. Finally, the strengths and limitations of the proposed approach are discussed, as well as promising avenues for future research directions.

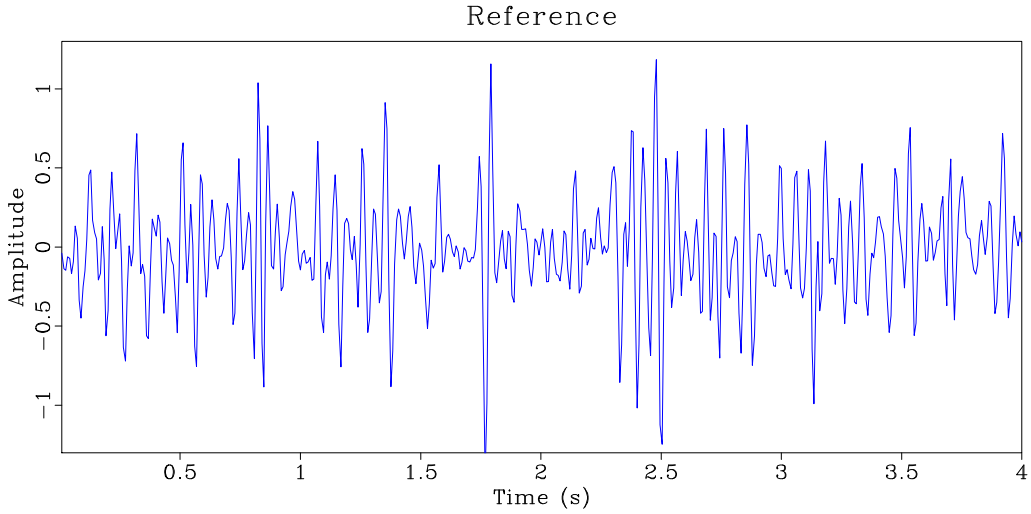
THEORY

I apply the concept of dynamic time warping (DTW) to flatten residual elliptical HTI moveout in image gathers generated using a processing workflow based on a VTI traveltime approximation that did not consider azimuthal anisotropy. Ideally, the processing workflow would account for HTI anisotropy. However, that process is computationally expensive, motivating the approximation method presented here. I begin by first stacking an image gather by computing the average trace value. Misalignment of traces in the initial gather acts as a low pass filter, so this stack provides the low frequency gather information. For each trace in the gather I use DTW to determine the shifts that best align that trace with the initial low frequency gather stack, thus focusing events. This generates a function which may be used to warp constituent traces to correct for anisotropy and flatten the gather. Flattened gathers may then be stacked to create an enhanced image.

My gather flattening workflow is based on the use of DTW to match pairs of seismic signals, so I first illustrate the function of DTW by reproducing an example featured in Hale (2013). That paper provides a mathematical treatment of the DTW process as well as useful pseudocodes and advice for a successful implementation. I begin with a synthetic reflectivity section containing 501 samples of 8 ms and apply

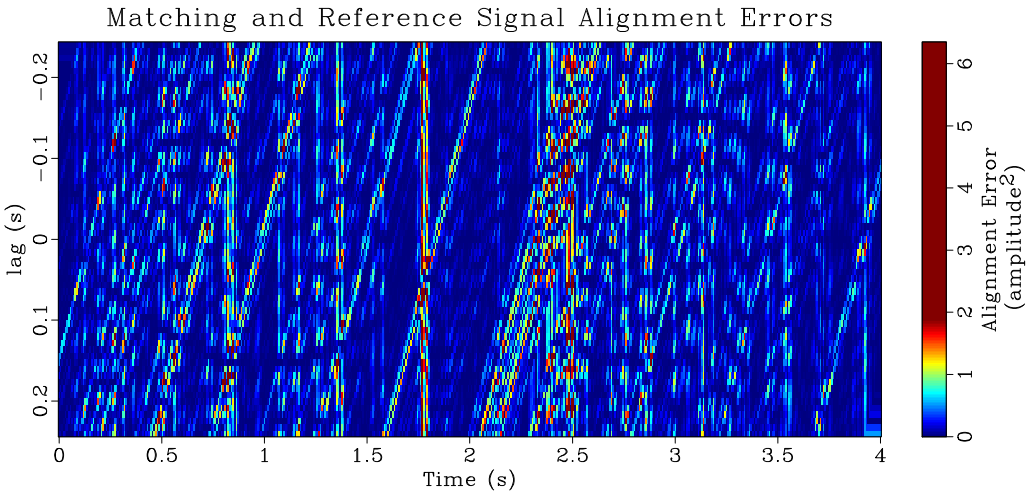


(a)

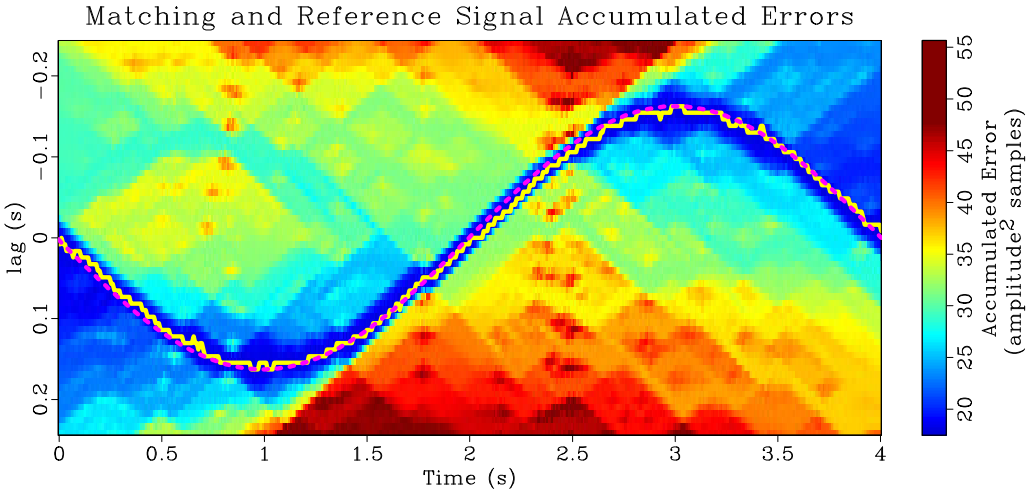


(b)

Figure 2.1: Input (a) matching and (b) reference traces used in DTW illustration reproduced from Hale (2013). Both traces are generated by convolving a reflectivity model with a 20 Hz Ricker wavelet and adding bandpassed noise. Applying a set of sinusoidal shifts with a maximum amplitude of 20 samples to the reflectivity used to generate the matching trace creates the reflectivity used to generate the reference trace.



(a)



(b)

Figure 2.2: DTW illustration reproduced from Hale (2013): (a) Alignment errors resulting from the difference between the reference and matching trace when the reference trace is shifted by various lags; (b) Accumulated errors resulting from a symmetric error accumulation over the top panel subject to strain limitations. Solid yellow line plots the DTW shifts calculated through strain limited backtracking of the accumulated errors. Dotted fuchsia line plots the ideal shifts.

sinusoidal shifts with a maximum amplitude of 20 samples to generate a shifted reflectivity model. Both reflectivity models are convolved with a 20 Hz Ricker wavelet and then bandpassed noise is added to create two synthetic traces. The trace resulting from the convolution of the wavelet with the initial reflectivity model is called the matching trace and is shown in Figure 2.1(a). The trace resulting from the convolution of the wavelet with the shifted reflectivity is called the reference trace and is shown in Figure 2.1(b).

Alignment errors, shown in Figure 2.2(a), are determined by computing the difference between the matching and reference signals at each point in the traces after shifting the matching trace by a set of lags, shown as the vertical coordinate. The sinusoidal path through the alignment errors panel defining the shifts is visible but difficult to follow. To make the path more discernible I accumulate over the alignment errors, shown in Figure 2.2(b). This involves starting on the left of the alignment errors panel, and for each lag in the initial alignment errors trace, selecting the lag in the next trace sample with minimum error subject to a strain limitation of 0.2, meaning that the shifts may change by a maximum of one sample over five trace samples. The minimal permissible error of that next trace is added to the error of the previous lag, leading to the term "accumulation". These accumulated errors are written to the lag index in that subsequent trace, and the process is repeated, moving across the alignment errors panel to create the accumulated errors panel, Figure 2.2(b). To remove the bias of general increase to the right in the accumulated errors panel, I accumulate from left to right and then from right to left. Accumulation both smooths the alignment errors and makes the optimal path more apparent,

forming a blue "valley".

To compute the minimizing shifts I perform backtracking. To begin, I select the shift corresponding to the lag with the minimal accumulated error in the far right trace of Figure 2.2(b). With this shift established I work backwards, picking the preceding permissible shift subject to the established strain limitations whose accumulated error value is minimal. This generates the set of shifts plotted in yellow in Figure 2.2(b). Ideal shifts are plotted in dashed fuchsia, and have good agreement with the calculated shifts. Applying the calculated shifts to the matching trace will warp its values so that they match the reference trace as closely as permissible.

Returning to the gather domain, suppose I average the traces within a gather to make a stack and then use DTW to match each constituent trace within that gather to the stack. This matching operation provides a set of shifts for each trace which best align that trace to the stack. Note that although only integer shifts are considered, these shifts may vary with the trace sample index. In the presence of a HTI media, the residual moveout within a gather, and hence the flattening shifts, may be considered as periodic over $\cos(2(\theta - \beta))$ where θ is gather azimuth and β is the anisotropic fast-axis orientation (Mallick et al., 1997; Grechka and Tsvankin, 1998). I note that if my migration does not take this type of anisotropy into account, traces in a gather along the fast-axis direction will be migrated with too low a velocity, and thus appear at a smaller time or shallower depth value than the stack, which represents an average trace for the gather. Similarly, traces along the slow axis direction will be migrated with too high a velocity and will appear at a greater time or depth than the stack. Based on how the shifts are defined in DTW, negative shifts "push" data downward

in the positive time or depth direction, while positive shifts "pull" it upward in the negative time or depth direction. Thus, the most negative shift will be aligned with the fast-axis of anisotropy and the most positive shift will be aligned with the slow anisotropic axis. If I assume that all other velocity and anisotropy corrections have been completed successfully, leaving only the component of moveout associated with elliptical HTI anisotropy, and that the anisotropic dependence on inclination ϕ is independent of azimuth, I may model my gather shifts as

$$\text{shifts}\left(\theta, \phi, \beta\right) = -\cos\left(2\left(\theta - \beta\right)\right)\ R\left(\phi\right), \tag{2.1}$$

where $R(\phi) \geq 0$ describes the dependence of the shift amplitude with inclination ϕ. Therefore, by determining the phase orientation β of the shifts, I also determine the anisotropic fast-axis orientation. In materials where anisotropy is caused by coherent fracture networks, the fast-axis will be the same as the primary fracture network orientation, and thus the principal stress direction.

I solve for phase argument β by integrating the test functions $u\left(\theta, \alpha\right)$ against the shifts to interrogate the data for the correct phase argument. If I chose my test functions to have the form

$$u\left(\theta, \alpha\right) = -\cos\left(2\left(\theta - \alpha\right)\right), \tag{2.2}$$

then I may write

$$\beta = \underset{\alpha \in [0,180)}{\arg\max} \frac{\int\int u\left(\theta, \alpha\right)\ \text{shifts}\left(\theta, \phi, \beta\right)\ d\theta\, d\phi}{\int\int u\left(\theta, \alpha\right)^2\ d\theta\, d\phi}, \tag{2.3}$$

where β, the orientation of the principal axis of anisotropy, is equal to the α which maximizes equation 2.3. The difference between the maximizing and minimizing

values of this integral describes the relative strength of the anisotropy, generating an attribute I call anisotropic intensity. A mathematical treatment of anisotropic intensity is now presented.

Suppose I have a collection of gather shifts modeled by equation 2.1,

$$\text{shifts}\left(\theta,\phi,\beta\right) = -\cos\left(2\left(\theta-\beta\right)\right)\ R\left(\phi\right), \tag{2.4}$$

defined for $\theta \in [0, 360)$, $\phi \in [\phi_o, \phi_f]$, $\phi_f > \phi_o \geq 0$, where θ describes the azimuth and ϕ the inclination within a gather. I also require that $R\left(\phi\right) \geq 0$. This model implies that the shift dependance on azimuth and offset may be separated, and that changes in inclination will not result in shifts reversing polarity. Suppose I want to determine the value of some fixed $\bar{\beta} \in [0, 180)$, a phase shift aligned with the orientation of the principal axis. I may introduce a family of test functions according to equation 2.2,

$$u\left(\theta,\alpha\right) = -\cos\left(2\left(\theta-\alpha\right)\right), \tag{2.5}$$

featuring a phase shift parameter $\alpha \in [0, 180)$. These test functions are defined over the domain of the function $\text{shifts}\left(\theta,\phi\right)$. I wish to prove equation 2.3, so I may write my hypothesis as:

$$\Lambda(\alpha) = \frac{\int_{\phi_o}^{\phi_f}\int_0^{360} u\left(\theta,\alpha\right)\ \text{shifts}\left(\theta,\phi,\bar{\beta}\right)\ d\phi\, d\theta}{\int_{\phi_o}^{\phi_f}\int_0^{360} u\left(\theta,\alpha\right)^2\ d\theta\, d\phi} \quad \text{is maximized by } \alpha = \bar{\beta}. \tag{2.6}$$

I begin by substituting equations 2.4 and 2.5 into equation 2.6:

$$\Lambda(\alpha) = \frac{\int_{\phi_o}^{\phi_f}\int_0^{360} \cos\left(2\left(\theta-\alpha\right)\right)\ \cos\left(2\left(\theta-\bar{\beta}\right)\right)\ R\left(\phi\right)\ d\phi\, d\theta}{\int_{\phi_o}^{\phi_f}\int_0^{360} \cos^2\left(2\left(\theta-\alpha\right)\right)\ d\theta\, d\phi}. \tag{2.7}$$

If I let $\psi = 2\left(\theta - \alpha\right)$, $\psi_o = -2\alpha$, $\psi_f = 720 - 2\alpha$, and through the linearity of integration I have:

$$\Lambda(\alpha) = \frac{\int_{\phi_o}^{\phi_f} R\left(\phi\right) d\phi}{\int_{\phi_o}^{\phi_f} d\phi} \frac{\int_{\psi_o}^{\psi_f} \cos\left(\psi\right) \cos\left(\psi + 2\left(\alpha - \bar{\beta}\right)\right) d\psi}{\int_{\psi_o}^{\psi_f} \cos^2\left(\psi\right) d\psi}. \tag{2.8}$$

Allowing $\Phi = \frac{\int_{\phi_o}^{\phi_f} R(\phi) d\phi}{\int_{\phi_o}^{\phi_f} d\phi}$, $\gamma = 2\left(\alpha - \bar{\beta}\right)$, $\Psi = \int_{\psi_o}^{\psi_f} \cos^2\left(\psi\right) d\psi$, and using the identity $\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ provides:

$$\hat{\Lambda}(\gamma) = \frac{\Phi}{\Psi}\left(\cos(\gamma)\Psi - \sin(\gamma)\int_{\psi_o}^{\psi_f} \cos(\psi)\sin(\psi)\, d\psi\right). \tag{2.9}$$

I note that $\Psi > 0$ because $\cos^2(\psi) > 0$ except on a set of measure zero, and $\psi_f > \psi_o$. Furthermore, $\Phi \geq 0$ since $R\left(\phi\right) \geq 0$ and $\phi_f > \phi_o$. $\Phi = 0$ implies that $R\left(\phi\right) = 0\ \forall\phi$, or that there is no azimuthal anisotropy observable over the ray path at that time or depth in the gather. In that case the concept of a principal anisotropic axis is meaningless and $\hat{\Lambda}(\gamma) = 0\ \forall\gamma$, so I will disregard it and assume $\Phi > 0$, or that there is some observable azimuthal anisotropy present. Furthermore, $\int_{\psi_o}^{\psi_f} \cos(\psi)\sin(\psi)\, d\psi = 0$ since:

$$\int_{\psi_o}^{\psi_f} \cos(\psi)\sin(\psi) d\psi = -\int_{\cos(\psi_o)}^{\cos(\psi_f)} \xi d\xi \tag{2.10}$$

and $\cos(\psi_o) = \cos(\psi_o + 720) = \cos(\psi_f)$. Therefore:

$$\hat{\Lambda}(\gamma) = \Phi\cos(\gamma). \tag{2.11}$$

To determine the γ that maximizes $\hat{\Lambda}$ I take the first and second derivatives:

$$\frac{d\hat{\Lambda}}{d\gamma} = -\Phi\sin\left(\gamma\right) \tag{2.12}$$

and

$$\frac{d^2\hat{\Lambda}}{d\gamma^2} = -\Phi\cos(\gamma)\,. \tag{2.13}$$

Equation 2.11 achieves a maxima where equation 2.12 is zero and equation 2.13 is negative. This occurs whenever γ is a multiple of 360. Due to restrictions imposed on α and $\bar{\beta}$, $\gamma \in (-180, 180)$, so the only permissible maximizing value within that interval is $\gamma = 0$ or equivalently $\alpha = \bar{\beta}$. To confirm that this is indeed the maximizing value, note the limit of the second derivative of $\hat{\Lambda}(\gamma)$ in equation 2.13 as $\gamma \to \pm 180$ is positive, indicating that at the edges of the domain $\hat{\Lambda}(\gamma)$ approaches a minima rather than a maxima.

Because the test functions $u(\theta, \alpha)$ have no dependance on ϕ, it is trivial to show the same result holds for image gathers that are sampled in azimuth and offset rather than azimuth and inclination.

This derivation allows me to provide an explicit definition of the anisotropic intensity attribute. I defined anisotropic intensity to be the difference between the maximizing and minimizing value of equation 2.3. Examining the derivation in this appendix, I see that the maximizing value occurs at $\alpha = \bar{\beta}$. The minimizing value occurs when equation 2.12 is equal to 0 and equation 2.13 is positive, which happens in the limit of $\gamma \to \pm 180$. This limit is equivalent to $\alpha \to \bar{\beta} \pm 90$, with the addition or subtraction used for finding α defined with modulo 180 so it "wraps" from 180 to 0. This result is intuitive, as I expect the anisotropic slow axis, where the minimum is attained, to be perpendicular to the anisotropic fast-axis. Thus, if I assume γ attains

the minimizing value, I may define anisotropic intensity Υ:

$$\Upsilon = 2\Phi, \tag{2.14}$$

or equivalently:

$$\Upsilon = 2\frac{\int_{\phi_o}^{\phi_f} R\left(\phi\right) d\phi}{\int_{\phi_o}^{\phi_f} d\phi}, \tag{2.15}$$

so Υ is a measure of the average amplitude of "wiggle" caused by the anisotropic ellipse across the gather. $\Upsilon = 0$ implies that no azimuthal anisotropy is present over the ray path. Increasing the average "wiggle" increases the value of Υ.

This method works for gathers with arbitrary sampling in ϕ and θ, provided sufficient distribution of θ samples to avoid aliasing. Based on the way the method was constructed, it also works for gathers sampled in offset rather than inclination, and traces sampled in either depth or time. If the shifts have a strong positive or negative bias it may be beneficial to correct them so the summation of shifts over the trace index for each time or depth within a gather is equal to zero.

SYNTHETIC GATHER EXAMPLE

I use a synthetic example to demonstrate my method. Figure 2.3(a) contains an input synthetic spiral gather designed to simulate elliptical HTI anisotropy whose orientation and intensity vary with time. This gather is constructed by first taking a synthetic trace and spraying that trace into an "ideal" flat gather. Elliptical HTI effects are simulated by using normal moveout (NMO) modeling to generate a set of common azimuth gathers. Each of these gathers is modeled using a constant gradient

velocity profile multiplied by an elliptic azimuthal anisotropy component in the form of Mallick et al. (1997) whose orientation varies with time. NMO correction using the constant gradient velocity profile is applied to each of the common azimuth gathers, leaving residual moveout from the unaccounted anisotropic component of the velocity profile. The spiral gather in Figure 2.3(a) is constructed from these common azimuth gathers by selecting appropriate offset and azimuth pairs. Although this spiral gather was constructed by applying isotropic NMO corrections to simulate an elliptical HTI medium, it effectively approximates HTI moveout in an image gather.

Notice that the periodicity of the elliptical HTI "wiggles" in the gather change with trace index as a result of irregular sampling of azimuth caused by the gather's spiral structure.

The input gather is stacked to create a reference trace, and DTW performed to determine the shifts which match each of the traces within the gather to the stack. To avoid cycle skipping, a strain limit is placed on the shifts so that they are unable to change too rapidly with time. These shifts are shown in Figure 2.3(b). The blue bands, indicating the most negative shifts, correspond to the anisotropic principal, or fast, axis. The red bands indicate the anisotropic slow axis. The bands slope downward to the right, indicating the change of anisotropic axis orientation over time. The shifts in Figure 2.3(b) are applied to the traces in Figure 2.3(a) to generate the flattened gather shown in Figure 2.3(c).

To determine the orientation of the principal axis, I generate a suite of test functions according to equation 2.2 and integrate them against the gather shifts as

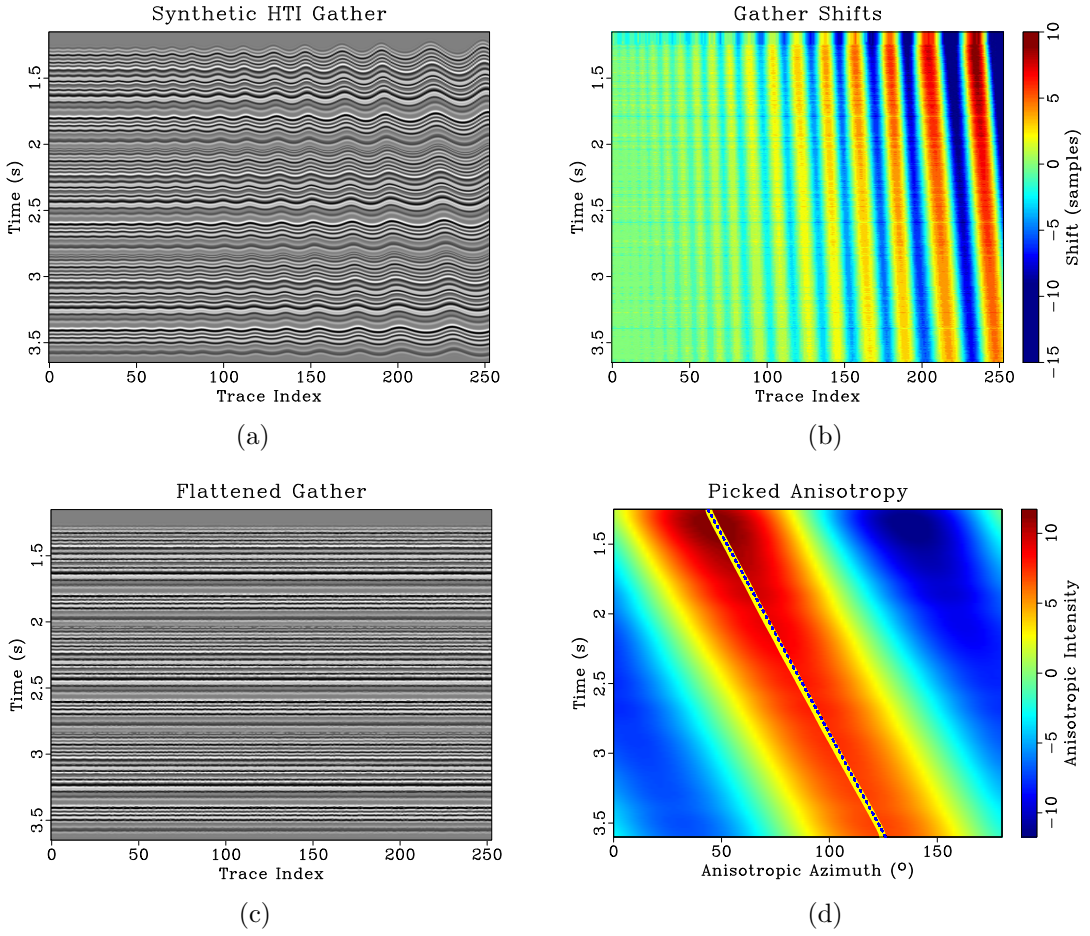


Figure 2.3: Synthetic example of dynamic time warping for gather flattening and determining principal anisotropic axis: (a) Synthetic gather with residual elliptical anisotropy; (b) Shifts that map the traces of Figure 2.3(a) to its stack; (c) Flattened gather resulting from applying the shifts in Figure 2.3(b) to the traces in Figure 2.3(a); (d) Determining the principal HTI axis based on the shifts in Figure 2.3(b). Background plots the value of equation 2.3 for different anisotropic azimuths. Solid yellow line plots the picked anisotropic azimuth which maximizes that equation at each time. Dotted blue line plots the ideal anisotropic azimuth. The difference between the maximizing and minimizing value of equation 2.3 at each time becomes the anisotropic intensity.

described by equation 2.3. The result of this integral is plotted as the background color in Figure 2.3(d). The maximum value of this integral is automatically picked at different times using the method described in Fomel (2009b) and plotted as a solid yellow line in Figure 2.3(d). This value provides β, the anisotropic azimuth. The difference between the underlying value maximizing the integral and the minimizing value provides anisotropic intensity, a measure of the anisotropic "wiggle" size over the gather at that time. The ideal anisotropic azimuth is plotted as a dashed blue line in Figure 2.3(d) which overlays the picked anisotropy, indicating that the method has successfully recovered its value.

I generate a series of stacks to illustrate the frequency content uplift resulting from gather flattening. Figure 2.4(a) contains a stack of the input gather from Figure 2.3(a) featuring residual elliptical HTI moveout. This stack is what traces are matched to during the dynamic warping process. Figure 2.4(b) plots the stack of the flattened gather from Figure 2.3(c). Notice that this stack has higher frequency content than the input stack, where the residual elliptical HTI anisotropy has acted as a low-pass filter. For comparison, the stack of the ideal gather is shown in Figure 2.4(c). This stack closely resembles that of the flattened gather, indicating that the gather flattening process has recovered high-frequency information missing from the input gather's stack.

FIELD DATA EXAMPLE

I show how my method is able to enhance seismic images and improve their bandwidth by applying it to a 3D field dataset that was depth imaged using Kirchhoff migration

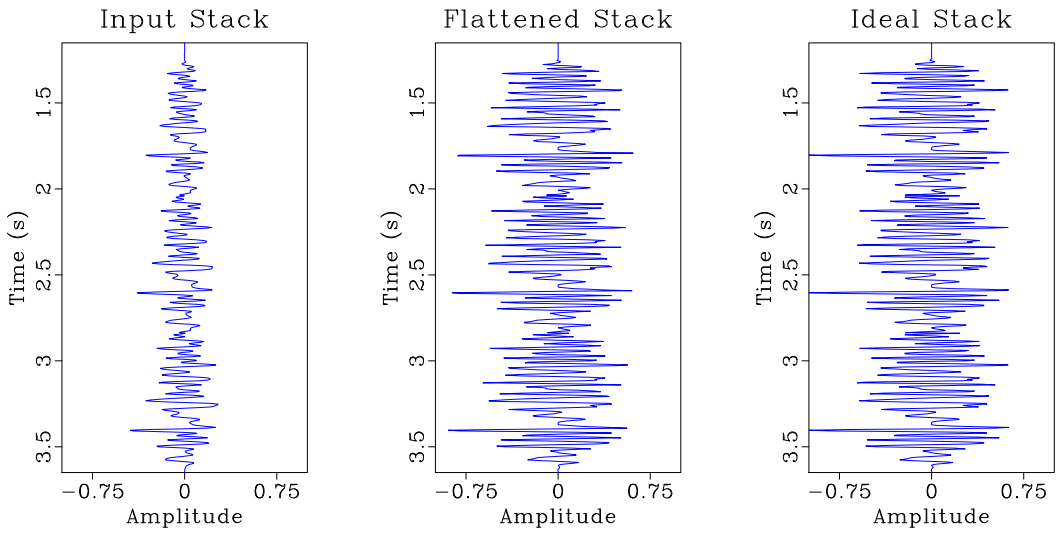


Figure 2.4: Stacks corresponding to: (a) the input gather in Figure 2.3(a) with residual elliptical HTI anisotropy; (b) the flattened gather in Figure 2.3(c); (c) the ideal flat gather.

with a VTI traveltime approximation. The migration velocity field was determined using VTI tomography.

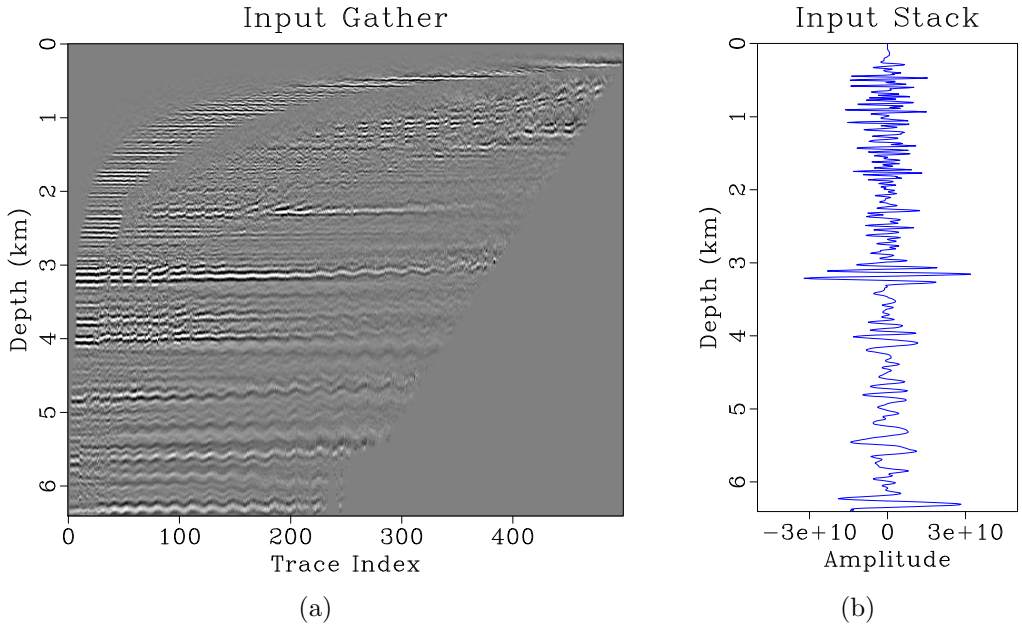


Figure 2.5: Gather flattening inputs: (a) common image gather with spiral structure; (b) the gather stack.

I begin by flattening the gathers within the seismic volume by applying the same workflow as in my synthetic example. Shifts that best match traces within each gather to its stack are calculated and then applied. Flattened gathers are stacked to generate an enhanced seismic image, called the flattened stack, and the shifts are integrated against a set of test functions according to equation 2.3 to determine the principal anisotropic axis and anisotropic intensity.

Figure 2.5(a) contains an example of a spiral gather from this survey, and Figure 2.5(b) plots its stack, or average trace value. Energy in the gather appears to

have a periodic wiggle across the trace indexes. The periodic moveout is caused by horizontally transverse anisotropy, where the seismic velocity is faster in one azimuthal direction than the other. Some moveout with offset is also present, visible as a general trend upward to the right in some of the events within the gather. This residual moveout is due to the tomography process, where the VTI velocity model is initially generated on a fine grid, and then smoothed on a coarse grid. The smoothed coarse model is used to migrate data, which results in near- and mid-offset traces being properly aligned in the gather at the expense of some far-offset traces.

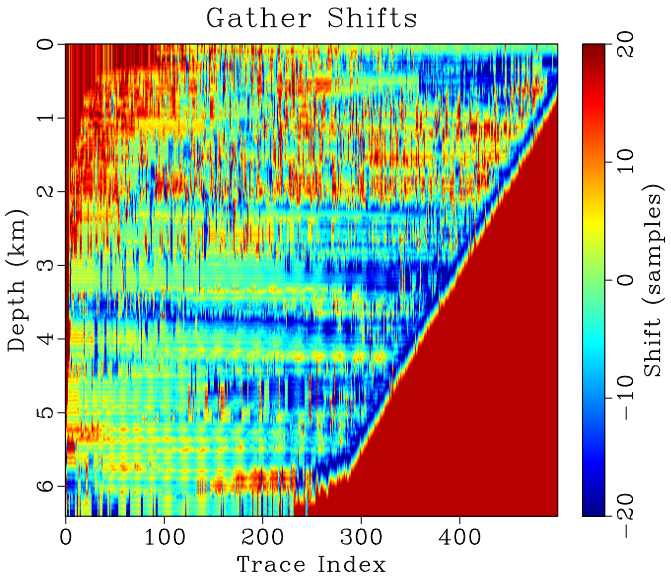


Figure 2.6: The shifts that most closely match traces within the input gather in Figure 2.5(a) to the stack shown in Figure 2.5(b).

For each trace, I determine the shifts that best match that trace to the stack. These shifts are shown in Figure 2.6. Shifts are applied to traces in the input gather to create the flattened gather in Figure 2.7(a) and its stack in Figure 2.7(b). Notice

that the gather flattening process has corrected for the residual moveout with offset present in far-offset traces as well as the periodic residual HTI moveout.

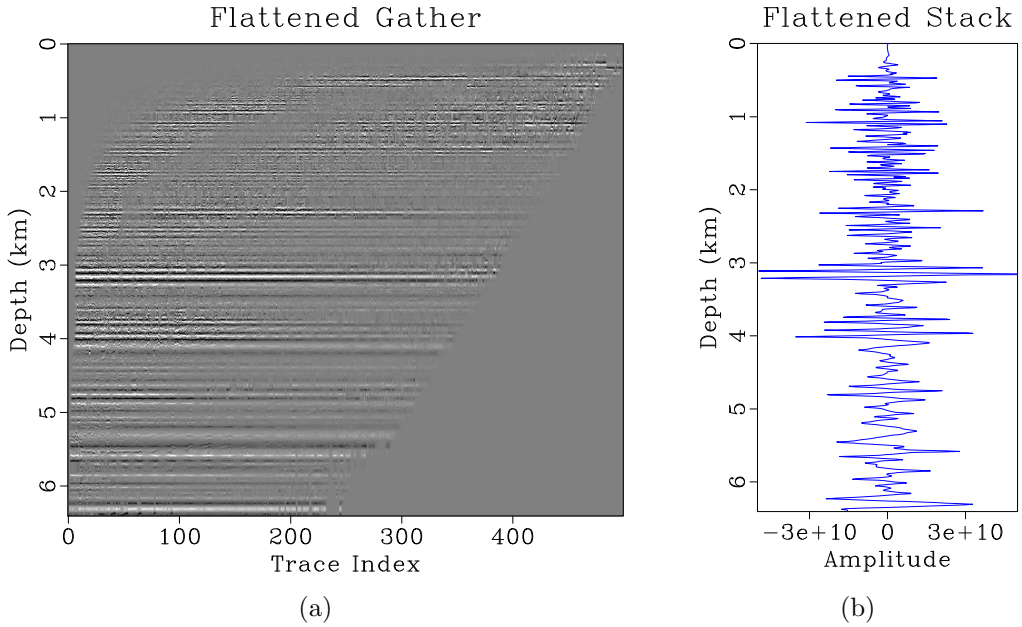


Figure 2.7: Gather flattening outputs: (a) flattened common image gather generated by applying the shifts in Figure 2.6 to the input gather in Figure 2.5(a); (b) the flattened gather stack.

Similar to the synthetic gather example, the azimuth of the principal axis is determined by finding the azimuthal anisotropy orientation β in equation 2.1 that best fits the periodic portion of gather shifts. Then, the anisotropic intensity is determined. This process is illustrated in Figure 2.8.

I repeat this process on all gathers within the seismic volume to enhance the stack, and determine anisotropic azimuth and intensity. Stacking in this example involves the average of all non-zero traces at each depth point within a gather, providing

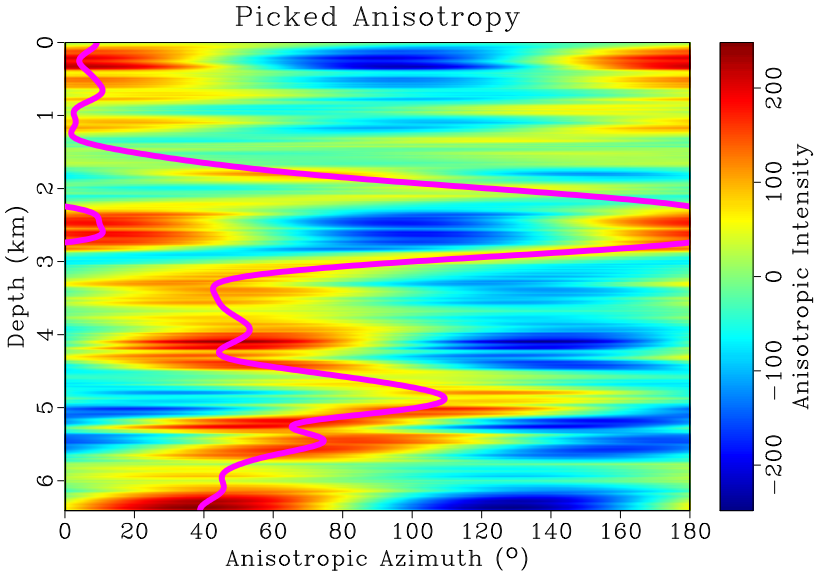


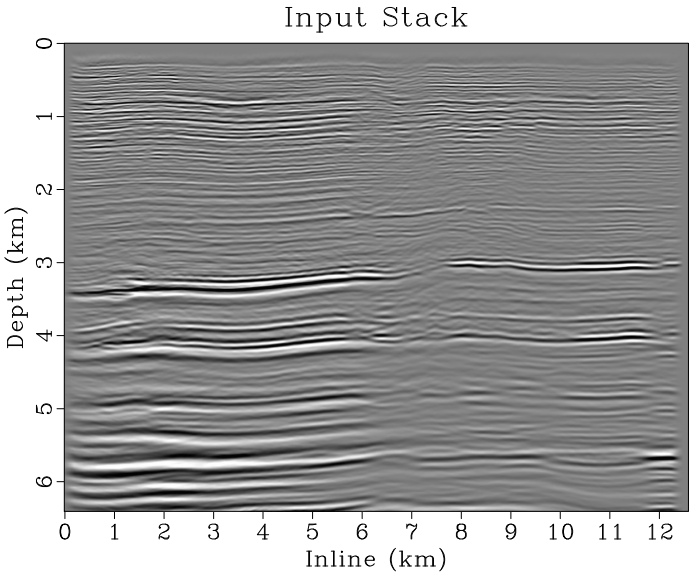
Figure 2.8: Illustration of determining the anisotropic parameters on this example gather based on the shifts in Figure 2.6. Background plots the value of equation 2.3 for different anisotropic azimuths. Solid fuchsia line plots the picked anisotropic azimuth which maximizes that equation at each time. The difference between the maximizing and minimizing value of equation 2.3 at each depth becomes the anisotropic intensity for this gather's spatial position.

equal weight to non-zero samples.

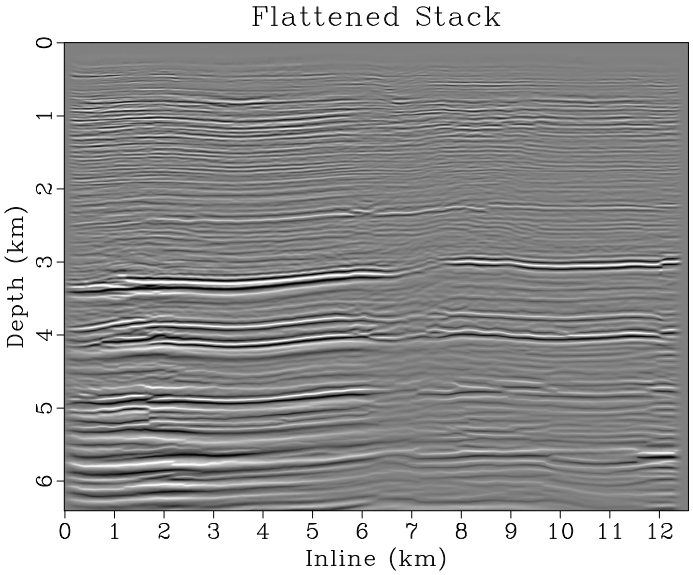
Figure 2.9(a) contains a constant crossline seismic image generated from stacking gathers prior to flattening. Figure 2.9(b) is the image that results from stacking the same gathers after flattening. Reflection events within the flattened stack generally appear sharper, more focused, and more coherent. Also note that events present in the flattened image have some corresponding signal in the input image – the process has not created new structure or shapes in the image.

Figure 2.10(a) contains a first zoomed panel from the lower left portion of Figure 2.9(a) and Figure 2.10(b) contains the corresponding zoomed flattened stack from Figure 2.9(b). Figures 2.10(a) and 2.10(b) contain zoomed inline slices of the input and flattened stacks, respectively. Notice how reflector discontinuities centered beneath 1.7 km are more apparent in the flattened stack. Reflection events appear sharper and more focused. Events in Figure 2.10(b) have some corresponding shapes in the image shown in Figure 2.10(a), illustrating how the method amplifies already present structures. Figure 2.10(c) contains the wavenumber spectra of the two zoomed panels. The blue line corresponds to the input stack in Figure 2.10(a) and the red line corresponds to the flattened stack in Figure 2.10(b). The red spectrum of the flattened stack contains more energy at higher wavenumbers than the blue spectrum of the input stack.

Figure 2.11(a) contains a second zoomed panel from the central portion of Figure 2.9(a) and Figure 2.11(b) contains the corresponding zoomed panel of the flattened stack from Figure 2.9(b). Events crossing the "washed out" portion of the



(a)



(b)

Figure 2.9: Constant crossline slices of seismic volume visualizing: (a) the input stack and (b) the flattened stack.

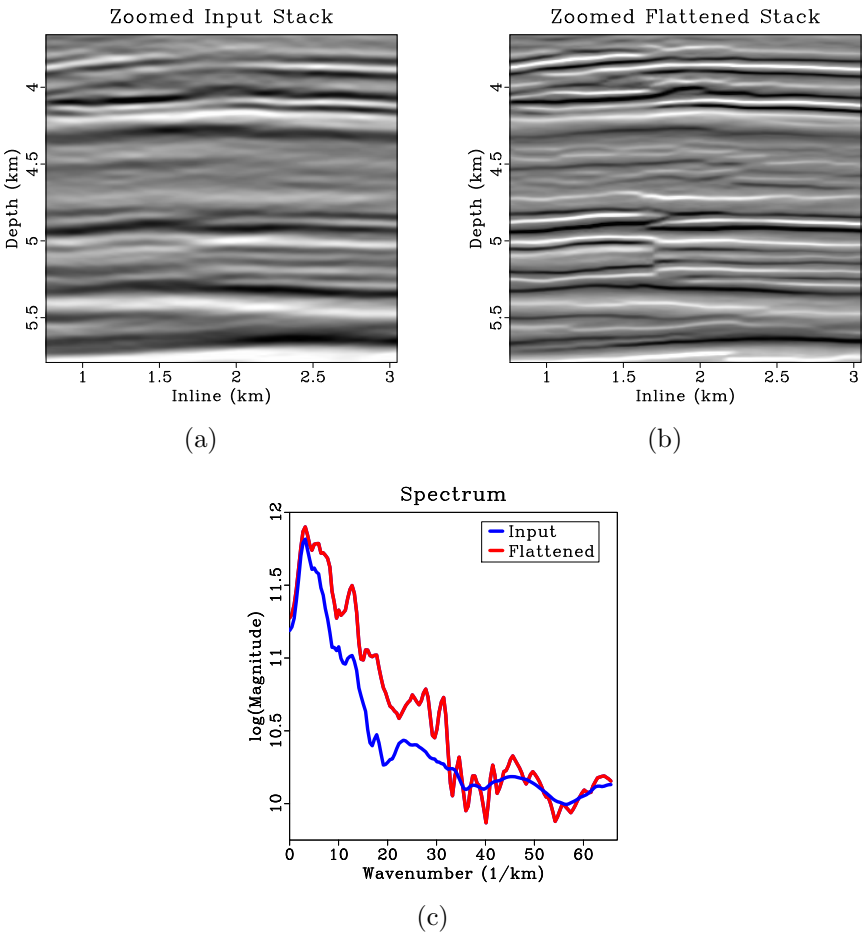


Figure 2.10: Zoomed inline slices corresponding to: (a) the input stack and (b) the flattened stack; (c) visualizes the wavenumber spectrum of Figure 2.10(a), the input stack, in blue and Figure 2.10(b), the flattened stack, in red.

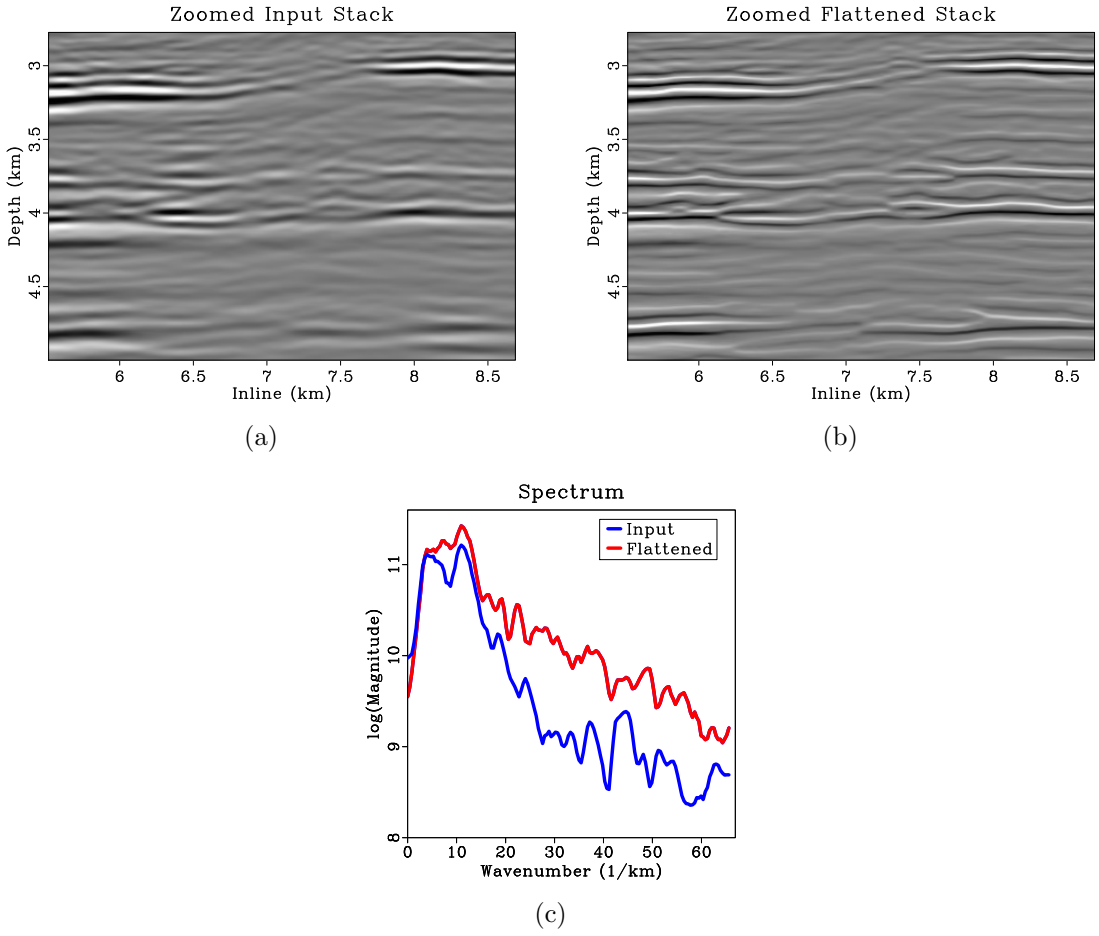
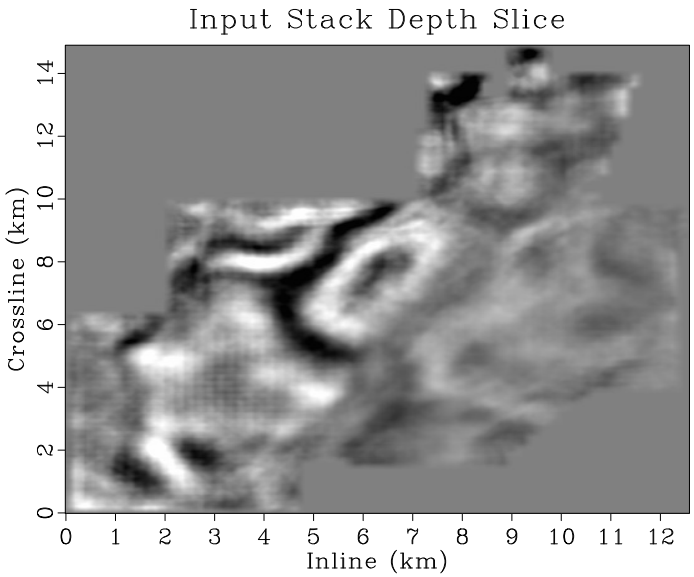


Figure 2.11: A second set of zoomed panels originating from Figures 2.9(a) and 2.9(b) corresponding to: (a) the input stack and (b) the flattened stack; (c) visualizes the wavenumber spectrum of Figure 2.11(a), the input stack, in blue and Figure 2.11(b), the flattened stack, in red.

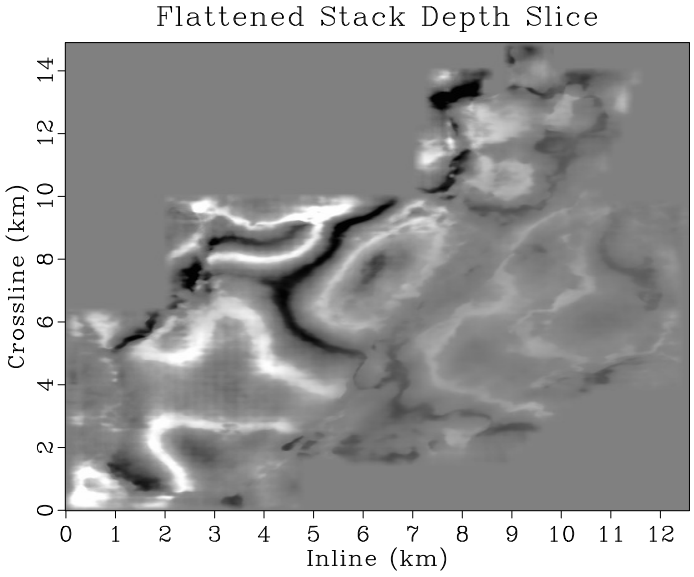
image between 6.7-7.6 km are more coherent in the flattened stack, and an automatic picking algorithm would likely have an easier time following these more coherent horizons across the washout. Figure 2.11(c) contains the wavenumber spectrum of these two zoom panels. The blue line corresponds to Figure 2.11(a) and the red line corresponds to Figure 2.11(b). As with Figure 2.10(c), the spectrum of the flattened stack has greater energy at larger wavenumbers relative to the spectrum of the input stack and has increased bandwidth.

I visualize a first set of depth slices of the input and flattened stacks in Figures 2.12(a) and 2.12(b), and zoomed sections of those slices in Figures 2.13(a) and 2.13(b). Events in the flattened stack depth slices are significantly more coherent and focused than in the input stack slices, and the discontinuity caused by a fault running upward and to the right beginning at inline 3.5 km, crossline 0 km in Figures 2.12(a) and 2.12(b) is more easily visible in the flattened stack than in the input. Examining close-ups of the slices in Figures 2.13(a) and 2.13(b), I see that features which appear as smudges or blurs in the input stack appear as coherent events in the flattened stack.

A second set of depth slices for input and flattened stacks are shown in Figures 2.14(a) and 2.14(b). Again, events in the depth slices of the flattened stack are more coherent and focused than those of the input slice. Notice that a checkerboard artifact corresponding to the acquisition footprint is present for both the input and flattened slices in the lower left corner, highlighted in the zoomed depth slices of Figures 2.15(a) and 2.15(b). This illustrates how the process of flattening does not create new shapes in the image, but rather aligns traces so they more closely match



(a)



(b)

Figure 2.12: First set of constant depth slices visualizing: (a) the input stack and (b) the flattened stack.

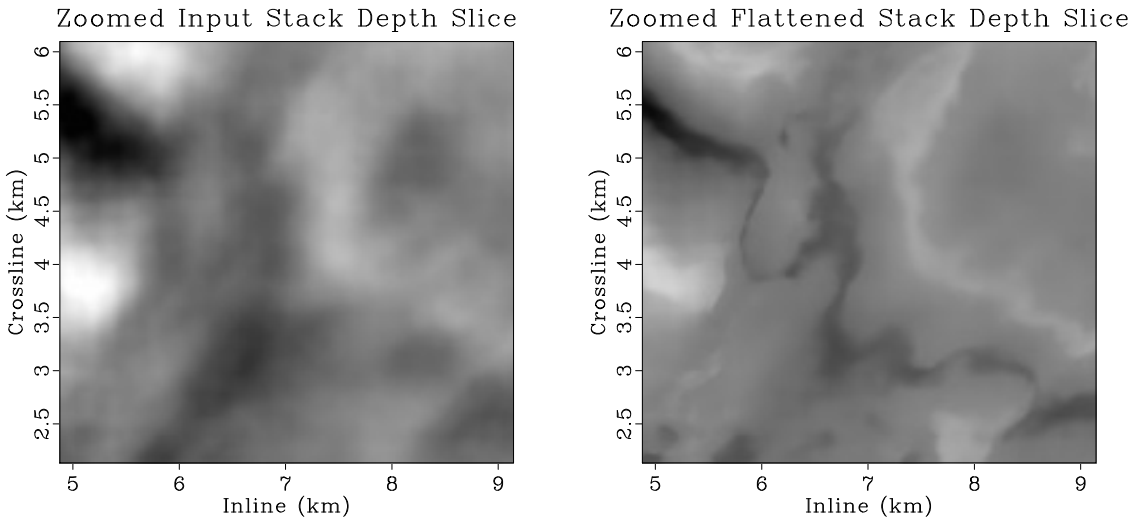
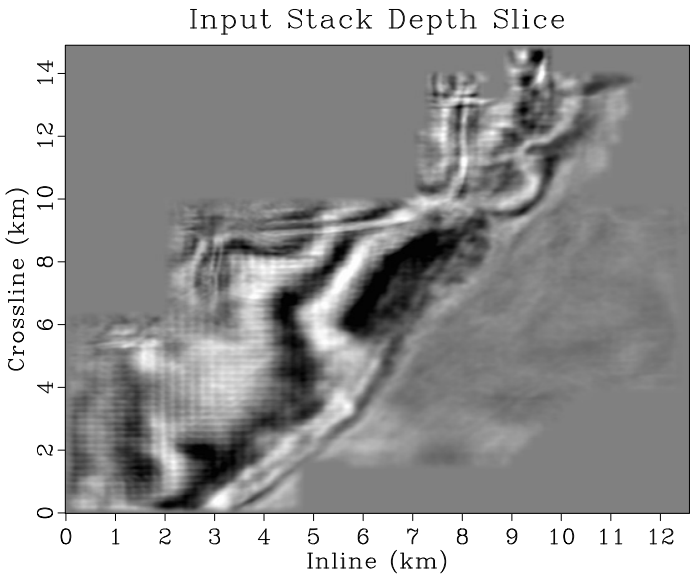


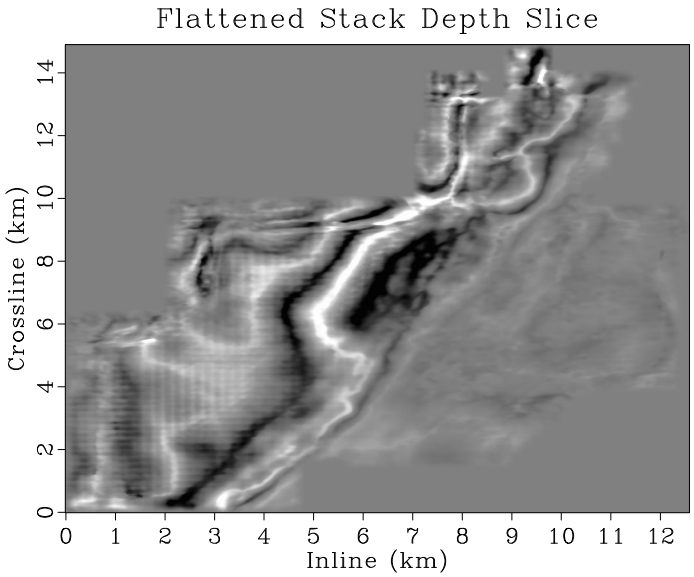
Figure 2.13: Zoomed panels originating from Figures 2.12(a) and 2.12(b) corresponding to: (a) the input stack and (b) the flattened stack.

existing shapes. The process of gather flattening will not suppress coherent artifacts that are present in an image, as it assumes any coherent feature is signal.

The anisotropic azimuth for the same depth slice as shown in Figures 2.14(a) and 2.14(b) is displayed in Figure 2.16(a), while Figure 2.16(b) contains the anisotropic intensity for that depth slice. For both of the attributes, the region between 0-6 km inline and 0-9 km crossline has relatively consistent values. Here, anisotropic azimuth values mostly range between $90 - 130°$, and anisotropic intensity values do not vary as greatly as in other portions of the depth slice. This appears to be a relatively homogeneous portion of the slice, with relatively consistent anisotropic orientation and intensity. Presumably, this implies that a relatively consistent fracture network orientation and distribution exists there. Other areas of the slice, where the azimuth

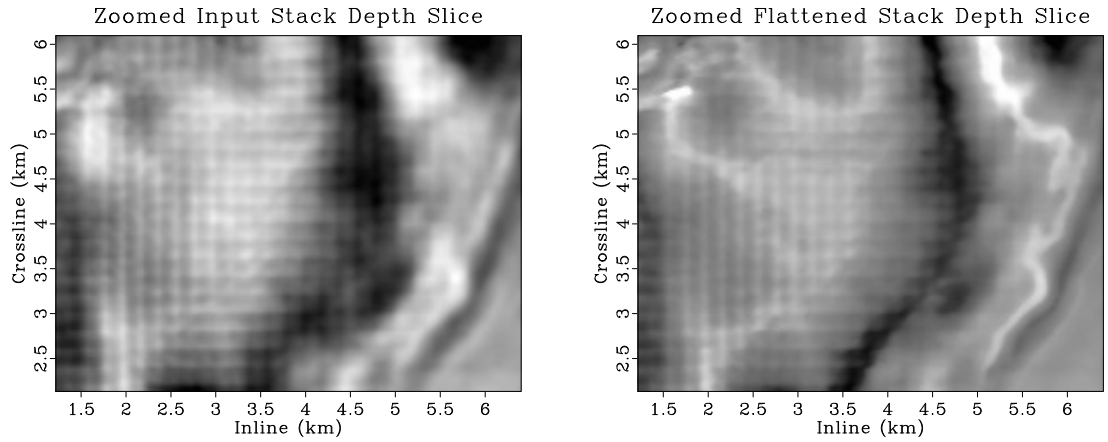


(a)



(b)

Figure 2.14: Second set of constant depth slices visualizing: (a) the input stack and (b) the flattened stack.



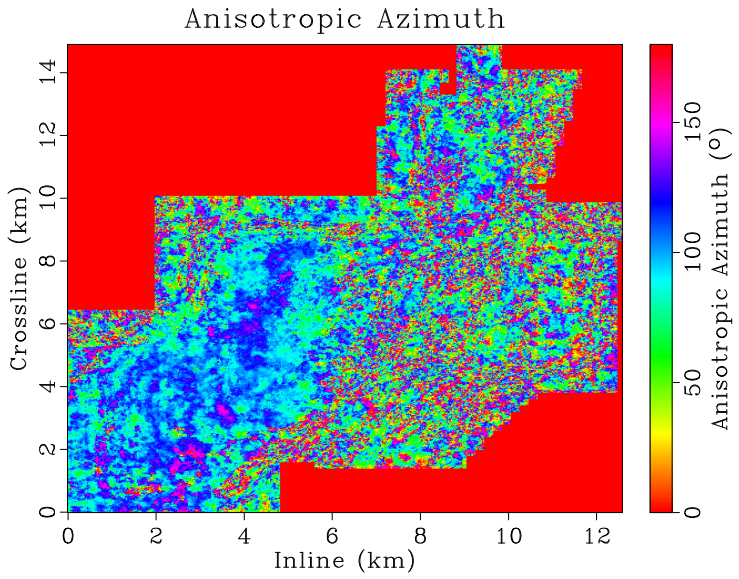
(a) (b)

Figure 2.15: Zoomed panels originating from Figures 2.14(a) and 2.14(b) corresponding to: (a) the input stack and (b) the flattened stack.

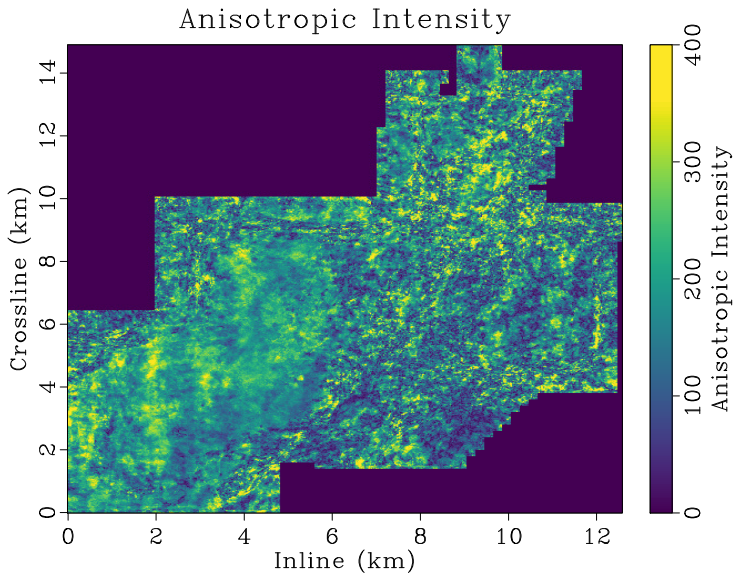
and intensity have greater variations, appear to indicate regions that may have a more chaotic distribution of fracture networks or other features causing elliptical HTI anisotropy. Results are consistent with proprietary log data acquired for wells in the study area.

DISCUSSION

The workflow proposed in this paper simultaneously generates enhanced seismic images and anisotropic attributes, thus leading to two different directions for further studies depending on if one is interested primarily in correcting for residual elliptical HTI anisotropy with dynamic warping or gaining additional information from the attributes generated.



(a)



(b)

Figure 2.16: Anisotropic attribute visualization from the same depth slice as in Figures 2.14(a) and 2.14(b): (a) the anisotropic azimuth, or the fastest azimuthal angle of seismic wave propagation, and (b) the anisotropic intensity or relative difference between the velocities of the fastest and slowest azimuthal directions.

The process of dynamic warping is simply an integer shift data matching problem, which makes its application computationally efficient. For my purposes it aligns one one-dimensional signal to another, seeking to make their values match as closely as possible given the constraints on rate of change and maximum shift provided. This may lead to several issues. First, moveout in the input gathers must already be "almost" correct. Even if the nearest peak or trough in the initial gather stack to one in the trace being matched does not actually correspond to the peak in that trace, it will still be matched to that feature. Hence, as much moveout correction as possible should be done prior to implementing this workflow.

Second, the method as described would not function well in the presence of a polarity reversal caused by an amplitude variation with offset (AVO) anomaly. Rather than matching a polarity reversed trough to a peak, it would simply match trough to trough. This issue could be overcome by first correcting for an AVO anomaly in the gathers, determining the shifts in those corrected gathers that would correct for elliptical HTI anisotropy, and then applying those shifts to gathers that have not had their AVO anomaly corrected. Different types of stack could also be used as the matching trace for gather flattening, or different portions of the gather could be matched to different stacks which are seen as most representative of traces within the interval to create a superior flattening result.

Finally, it is possible for seismic events to become distorted due to shifts applied by the warping algorithm. I did not observe this phenomenon in the field data experiment, but appropriately limiting the maximum strain and shift size parameters in the dynamic time warping algoritm is important for avoiding such distortions.

Although the method proposed in this paper focuses on applications to correct residual moveout related to elliptical HTI anisotropy, the data matching process could also shift traces to perform static correction. Residual moveout related to elliptical HTI anisotropy may be distinguished from that caused by static correction, which would appear as a constant shift applied to an entire trace. Note that because the DTW workflow treats each gather independently, static corrections computed by taking the average shift value over a trace would not necessarily be surface-consistent. If only a few traces within a gather feature residual moveout related to static correction, those constant shifts are unlikely to have a large affect on the HTI anisotropy attributes, as they would have to be periodic over 180° throughout the gather. If many traces feature moveout related to static correction, it would be beneficial to apply surface consistent correction to the seismic data, as this method for determining anisotropic axis and orientation assumes that the residual moveout is caused by HTI anisotropy.

Seismic images resulting from the stacking of flattened gathers contain more coherent and focused reflection events as shown in the constant crossline slices of Figures 2.9(a) and 2.9(b), their zoomed sections of Figures 2.10(a) and 2.10(b), and the depth slices of Figures 2.12(a) and 2.12(b). Flattening the gathers also overcomes the effective low-pass filter created by stacking gathers with residual elliptical HTI moveout. This is illustrated in the spectra of Figures 2.10(c) and 2.11(c), where the flattened stack spectra contain more energy at higher wavenumbers and greater bandwidth than the input stack spectra. Furthermore, this method does not create new structure or shape within the image, which is particularly obvious in Figures 2.15(a)

and 2.15(b). These more coherent events would likely be easier for an automatic interpretation or computer vision algorithm to follow. Therefore, the gather flattening and image enhancement process outlined here could fit well as part of an automatic interpretation workflow.

Another interesting direction for further study is investigation of the anisotropic azimuth and intensity parameters. These are functions of shifts which correct residual elliptical HTI moveout, and as such may be thought of as an average anisotropy measure over the entirety of a ray path, similar to how root-mean-square (RMS) velocity used in time migration is a measure of the average velocity over a ray path. The method assumes that anisotropy varies slowly in the subsurface, and measures the accumulated anisotropy along a ray path reflecting at a position in the subsurface. It approximates the subsurface elliptical HTI anisotropy field by highlighting areas where anisotropy is present, but may fail in regions where anisotropy values change rapidly. A useful extension could involve developing a transformation from the average anisotropy measures along a raypath which this method provides to a local or interval anisotropy. This local anisotropy could enable more accurate subsurface characterization, allowing for representations of local features rather than tendencies throughout the volume. A simple implementation of this could involve taking the derivative with respect to depth of a vector whose orientation and magnitude are defined by the anisotropic azimuth and intensity. A more complex version could involve a HTI ray tracing step and solving for the attributes throughout the volume based on the anisotropic attributes tied to those ray paths.

I have defined this paper's workflow so that each gather is independent, and

thus the processing may be ran in parallel, enabling a relatively simple implementation for large data sets. For the field data experiment in this paper, each gather has 500 traces and there are approximately 300,000 gathers within the volume. Running the process on 200 threads, the flattening of gathers and determination of principal axes was completed in under six hours, much faster than the time required for a processing workflow that took elliptical HTI anisotropy into account. I am not proposing a complete method for residual moveout correction, but rather a way of determining the orientation of the moveout whose correcting shifts may be modeled by an ellipse and a measure of how strong that elliptic component is. The method described in this paper is an inexpensive approximation to more costly anisotropic processing methods, but is not intended to replace them.

CONCLUSIONS

I propose a workflow that uses dynamic warping to efficiently correct for residual elliptical HTI moveout present in image gathers. The method works by calculating and applying the shifts that match each trace within an image gather to its stack. Stacking flattened gathers results in seismic images with more coherent and focused events. Fitting the shift dependence on azimuth to an ellipse provides both the azimuth of the fast anisotropic axis, which may coincide with the orientation of the primary fracture network, and a measure of the intensity of the anisotropy. These attributes may be used to aid in subsurface characterization. The method is embarrassingly gather-parallel, enabling it to be implemented relatively easily on large data sets, with much less computational expense than would be associated with a

processing workflow that takes the variation of seismic velocity with azimuth into account.

Chapter 3

Diffraction imaging and time-migration velocity analysis using oriented velocity continuation

I perform seismic diffraction imaging and time-migration velocity analysis by separating diffractions from specular reflections and decomposing them into slope components. I image slope components using migration velocity extrapolation in time-space-slope coordinates. The extrapolation is described by a convection-type partial differential equation and implemented in a highly parallel manner in the Fourier domain. Synthetic and field data experiments show that the proposed algorithms are able to detect accurate time-migration velocities by measuring the flatness

[1]Some of the material in this chapter was published as:

- Decker, L., and S. Fomel, 2014, Diffraction imaging and velocity analysis using oriented velocity continuation: 84th Annual International Meeting, SEG, Expanded Abstracts, 4810–4815. The authors contributed to this paper as follows: study conception and design: S. Fomel; data collection: L. Decker; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.
- Decker, L., 2014, Seismic diffraction imaging methods and applications: M.S. Thesis: The University of Texas at Austin. The author confirms sole responsibility for this thesis.
- Decker, L., D. Merzlikin, and S. Fomel, 2017a, Diffraction imaging and time-migration velocity analysis using oriented velocity continuation: Geophysics, **82**, no. 2, U25–U35. The authors contributed to this paper as follows: study conception and design: S. Fomel; data collection: L. Decker and D. Merzlikin; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.

of diffraction events in slope gathers for both single and multiple offset data.

INTRODUCTION

Seismic diffraction occurs when a seismic wave encounters a heterogeneity without a clearly defined tangent plane, such as an edge or tip, and the reflection part of the ray theory breaks down (Klem-Musatov, 1994). These divergent diffraction rays have similar behavior to a subsurface secondary source located at the heterogeneity (Keller, 1962). Analyzing diffraction moveout behavior in different domains can provide subsurface velocity information analogous to analyzing reflection moveout behavior from a surface source.

The fact that diffractions migrated with correct velocity collapse to points motivated Harlan et al. (1984) to propose the idea of separating diffractions from specular reflections and using diffraction focusing as a tool for velocity analysis. Separation of diffraction events from seismic data is a necessary step for velocity analysis because diffraction signals are typically significantly weaker than those of reflections (Klem-Musatov, 1994). Fomel et al. (2007) developed a constructive procedure for diffraction separation based on plane-wave destruction and diffraction focusing analysis based on velocity continuation and local kurtosis. The procedure was extended to 3-D azimuthally-anisotropic velocity analysis by Burnett and Fomel (2011). However, local kurtosis may not be an optimal measure for diffraction focusing because it requires smoothing or windowing in space, which reduces spatial velocity resolution through the smoothing window parameters.

A particularly convenient domain for separating diffractions and reflections and for analyzing migration velocities is dip-angle gathers (Brandsberg-Dahl et al., 2003; Biondi and Symes, 2004; Landa et al., 2008; Reshef and Landa, 2009; Klokov and Fomel, 2012). In the dip-angle domain, specular reflections appear as hyperbolic events centered at the reflector dip and bending upwards, even when over or under-migrated, and diffractions appear flat when imaged at the location of the diffractor with the correct velocity (Reshef, 2007). Measuring flatness of diffraction events in dip-angle gathers, as opposed to flatness of reflection and diffraction events in reflection-angle gathers, provides an alternative constraint on seismic velocity (Reshef and Landa, 2009). Traditionally, dip-angle gathers are constructed with Kirchhoff migration (Fomel and Prucha, 1999; Xu et al., 2001; Cheng et al., 2011; Koren and Ravve, 2011; Bashkardin et al., 2012; Klokov and Fomel, 2013).

In this paper, I adopt an analogous method to the dip-angle approach used by Reshef and Landa (2009) to devise a constructive and highly parallel procedure for estimating velocities in time-domain processing using data decomposition in slope (Ghosh and Fomel, 2012) and velocity continuation in the midpoint-time-slope domain. By analogy with the "oriented wave equation" (Fomel, 2003a), I call this approach *oriented velocity continuation* (OVC) and develop a fast spectral method for its implementation on common-offset data. This differs from the methods devised by Reshef and Landa (2009), which utilize a separate Kirchhoff-based angle prestack time or depth migration and calculation of travel time tables for each tested migration velocity. OVC uses a continuation approach where a single migration is used to determine an initial image in the midpoint-time-slope domain to which a veloc-

ity dependent phase shift is applied over the range of plausible migration velocities, enabling OVC to test a greater number of velocities at a lower computational cost.

Using a field-data experiment, I demonstrate the effectiveness of oriented velocity continuation in zero-offset diffraction imaging and velocity analysis, and using a synthetic model I observe that higher velocity resolution can be achieved when multiple offsets are included in the process.

ORIENTED VELOCITY CONTINUATION

Velocity continuation (Fomel, 2003b) is the imaginary process of a continuous transformation of seismic time-migrated images as they are propagated through different migration velocities. In the most general terms, the kinematics of velocity continuation can be described by an equation of the Hamilton-Jacobi type

$$\frac{\partial \tau}{\partial v} = F(v, \tau, \mathbf{x}, \nabla \tau) \ , \tag{3.1}$$

where $\tau(\mathbf{x}, v)$ is the location of a time-migrated reflector with time-domain coordinates $\mathbf{x} = (x_1, x_2, t)$ imaged with spatially constant time-migration velocity v. The particular form of function F in equation (3.1) depends on the acquisition geometry of the input data. For the case of common-offset 2D velocity continuation for data with half-offset h and slope p,

$$F(v, t, x, p) = v\, t\, p^2 \ + \frac{h^2}{v^3 t}, \tag{3.2}$$

and equation (3.1) corresponds to the characteristic equation of the image propagation process which describes a propagation of the time-migrated image $I(t, x, v)$ in

velocity v (Fomel, 2003b). Time-domain imaging can be performed effectively by extrapolating images in velocity and estimating velocity $v_m(t,x)$ of the best image (Larner and Beasley, 1987; Fomel, 2003b; Fomel and Landa, 2014a).

As shown by Fomel (2003a), it is possible to extend the formulation of a wave propagation process from the usual time-and-space coordinates to the phase space consisting of time, space, and slope. Applying a similar approach to equation (3.1), we first employ the Hamilton-Jacobi theory (Courant and Hilbert, 1989; Evans, 2010) to write the corresponding system of ordinary differential equations for the characteristics (velocity rays), as follows:

$$\frac{d\mathbf{x}}{dv} = -\nabla_p F \;, \tag{3.3}$$

$$\frac{d\mathbf{p}}{dv} = \nabla_x F + \frac{\partial F}{\partial t}\,\mathbf{p} \;, \tag{3.4}$$

$$\frac{dt}{dv} = F - \nabla_p F \cdot \mathbf{p} \;, \tag{3.5}$$

where $\mathbf{p}$ stands for $\nabla\tau$, the gradient or slope of time-migrated wavefield energy.

If the image $I(t,\mathbf{x},v)$ is decomposed in slope components $\widehat{I}(t,\mathbf{x},\mathbf{p},v)$ so that

$$I(t,\mathbf{x},v) = \int \widehat{I}(t,\mathbf{x},\mathbf{p},v)\,d\mathbf{p} \;, \tag{3.6}$$

I can then look for an equation that would adequately describe a continuous transformation of $\widehat{I}$. To preserve the geometry of the transformation, it is sufficient to require that $\widehat{I}$ transports along the characteristics described by equations (3.3-3.5). Applying partial derivatives and the chain rule, I arrive at the equation analogous to the Liouville equation (Engquist and Runborg, 2003):

$$\frac{\partial \widehat{I}}{\partial v} = (F - \nabla_p F \cdot \mathbf{p})\,\frac{\partial \widehat{I}}{\partial t} - \nabla_p F \cdot \nabla_x \widehat{I} + \left(\nabla_x F + \frac{\partial F}{\partial \tau}\,\mathbf{p}\right) \cdot \nabla_p \widehat{I} \;. \tag{3.7}$$

Equation (3.7) describes, in the most general form, the process of *oriented velocity continuation*, image propagation in velocity in the coordinates of time-space-slope. It is a linear first-order partial differential equation of convection type which operates in the phase space.

Common-offset oriented velocity continuation

To adopt the general theory described above to the case of common-offset 2D oriented velocity continuation, I can substitute equation (3.2) into (3.7), arriving at the equation

$$\frac{\partial \widehat{I}}{\partial v} = \left(\frac{h^2}{v^3 t} - \ v\,t\,p^2 \right) \frac{\partial \widehat{I}}{\partial t} - 2\,v\,t\,p\,\frac{\partial \widehat{I}}{\partial x} + v\,p^3\,\frac{\partial \widehat{I}}{\partial p}\ . \tag{3.8}$$

which describes image propagation in the time-space-slope coordinates rather than the usual time-space coordinates. After this kind of extrapolation, regular images can be reconstructed by stacking over offset and slope.

Slope gathers, analogous to dip-angle gathers, can be extracted before stacking over slope by analyzing $\{t, p\}$ panels for different image locations x and velocities v. Measuring flatness of diffraction events in these gathers provides a means for estimating migration velocity (Landa et al., 2008; Reshef and Landa, 2009).

For practical implementation, the formulation of oriented velocity continuation can be simplified by employing a stretch from the regular time coordinate to squared time $\sigma = t^2$ (Fomel, 2003b). According to this transformation, the Hamilton-Jacobi equation (3.1) becomes

$$\frac{\partial \sigma}{\partial v} = \frac{v}{2}\left(\frac{\partial \sigma}{\partial x}\right)^2 \ + \ 2\frac{h^2}{v^3}\ , \tag{3.9}$$

which leads to the simpler form of the oriented equation

$$\frac{\partial \widehat{I}}{\partial v} = \left(2\frac{h^2}{v^3} \ - \frac{v}{2}\, q^2 \right) \frac{\partial \widehat{I}}{\partial \sigma} - v\, q\, \frac{\partial \widehat{I}}{\partial x} \ . \tag{3.10}$$

where q corresponds to $\frac{\partial \sigma}{\partial x}$, and the image is constructed in $\{\sigma, x, q, h\}$ coordinates instead of $\{t, x, p, h\}$ coordinates. Applying the Fourier transform, we can further transform equation (3.10) to

$$\frac{\partial \tilde{I}}{\partial v} = i\,\omega \, \left(\frac{v}{2}\, q^2 \ - 2\frac{h^2}{v^3} \right) \tilde{I} - i\, v\, q\, k\, \tilde{I} \ , \tag{3.11}$$

where $\tilde{I}(\omega, k, q, v, h)$ is the double Fourier transform of $\widehat{I}(\sigma, x, q, v, h)$ in σ and x. Equation (3.11) has the analytical solution:

$$\tilde{I}(\omega, k, q, v, h) = \tilde{I}(\omega, k, q, v_0, h)\, e^{\, i\,(\omega\, q^2/4 + q\, k/2)\,(v^2 - v_0^2) \ + \ i\omega h^2 (\frac{1}{v^2} - \frac{1}{v_0^2})} \ , \tag{3.12}$$

where v_0 is a constant non-zero initial migration velocity.

Stacking over offset provides a slope-decomposed formulation for oriented velocity continuation:

$$\tilde{I}(\omega, k, q, v) = \sum_h \tilde{I}(\omega, k, q, v_0, h)\, e^{\, i\,(\omega\, q^2/4 + q\, k/2)\,(v^2 - v_0^2) \ + \ i\omega h^2 (\frac{1}{v^2} - \frac{1}{v_0^2})} \ . \tag{3.13}$$

This derivation suggests the following algorithm for time-domain imaging using common-offset 2D oriented velocity continuation:

1. Start with initial time migration with a constant velocity v_0 to generate $I(t, x, v_0, h)$.

2. Apply vertical time stretch to transform from t to σ.

3. Apply Fourier transform from σ to ω.

4. Perform slope decomposition (described in the next section) to generate $\hat{I}(\omega, x, q, v_0, h)$. Note that this operation is parallel in ω and h.

5. Apply Fourier transform from x to k to generate $\tilde{I}(\omega, k, q, v_0, h)$. Note that this operation is parallel in q and h.

6. Apply the phase-shift filter from equation (3.12) to generate $\tilde{I}(\omega, k, q, v, h)$ for multiple values of v. Note that this operation is data-intensive but parallel in q, k, and h.

7. Stack over offset to generate $\tilde{I}(\omega, k, q, v)$.

8. Apply inverse double Fourier transform to generate $\hat{I}(\sigma, x, q, v)$.

9. Apply inverse time stretch from σ to t.

10. Stack over q and extract the slice at time-migration velocity $v_m(t, x)$ to generate the final time-migrated image $I(t, x, v_m(t, x))$.

In order to estimate the velocity $v_m(t, x)$, I apply the workflow described above to diffraction imaging and modify it as follows:

- Before Step 1, I separate reflections and diffractions in the common-offset data using local plane-wave destruction (Fomel, 2002; Fomel et al., 2007; Decker et al., 2013).

- After Step 9, I analyze slope gathers $\hat{I}(t, x, q, v)$ and automatically pick the velocity $v_m(t, x)$ that corresponds to the maximum flatness (semblance) over q using the picking algorithm described by Fomel (2009b). This approach follows the principle of flatness of diffraction events in slope gathers (Landa et al., 2008; Reshef and Landa, 2009; Klokov and Fomel, 2012).

The computational cost associated with determining velocity using oriented velocity continuation is linear with the number of time samples, spatial samples, offsets, velocities, and slopes considered. It is parallel in spatial samples, offsets, velocity, and slope. The cost may then be considered as $O\left(\frac{N_t N_x N_h N_v N_p}{N_c}\right)$ where N_c is the number of cores available.

SLOPE DECOMPOSITION

In order to perform the initial slope decomposition (Step 4 in the algorithm above), I adopt the method of Ghosh and Fomel (2012). The idea of slope decomposition was discussed previously by Ottolini (1983) and implemented using the local-slant stack transform (Ventosa et al., 2012). The slope-decomposition algorithm suggested by Ghosh and Fomel (2012) is based on the time-frequency decomposition of Liu and Fomel (2013). Namely, at each frequency ω, I apply regularized non-stationary regression (Fomel, 2009a) to transform from space x to space-slope x-q domain. The non-stationary regression amounts to finding complex coefficients $A_n(\omega, x)$ in the decomposition

$$D(\omega, x) = \sum_{n=1}^{N_p} D_n(\omega, x) \ , \tag{3.14}$$

where $D(\omega, x)$ is the image slice, and $D_n(\omega, x)$ is its slope component corresponding to slope q_n:

$$D_n(\omega, x) = A_n(\omega, x)\, e^{i\,\omega\, x\, q_n}\ . \tag{3.15}$$

Equation (3.14) is the discrete analog of equation (3.6). Similarly to the time-frequency decomposition proposed by Liu and Fomel (2013), shaping regularization is used to control the variability of A_n coefficients and to accelerate the algorithm.

EXAMPLES

Toy Model Example

I first illustrate the concept of oriented velocity continuation using a simple toy model (Landa et al., 2008; Klokov and Fomel, 2013) with constant 1.0 km/s velocity containing one dipping and one flat reflector and a single diffractor centered at 0.5 km (Figure 3.1(a)). Figure 3.1(b) shows data warped to squared time.

Warped data are decomposed into their constituent slope components and initially under-migrated using $v_0 = 0.5\ \frac{km}{s}$. The initial slope decomposed image is shown in Figure 3.2, which illustrates a slope gather centered above the diffractor on the right panel, and a partial image containing energy with the slope of the top dipping reflector on the front panel. The partial image contains energy of the top dipping reflector, which has the selected slope, diffraction energy with that slope, and a small portion of the energy from the flat bottom reflector. Stacking over all constituent slopes provides an image (top left panel of Figure 3.4).

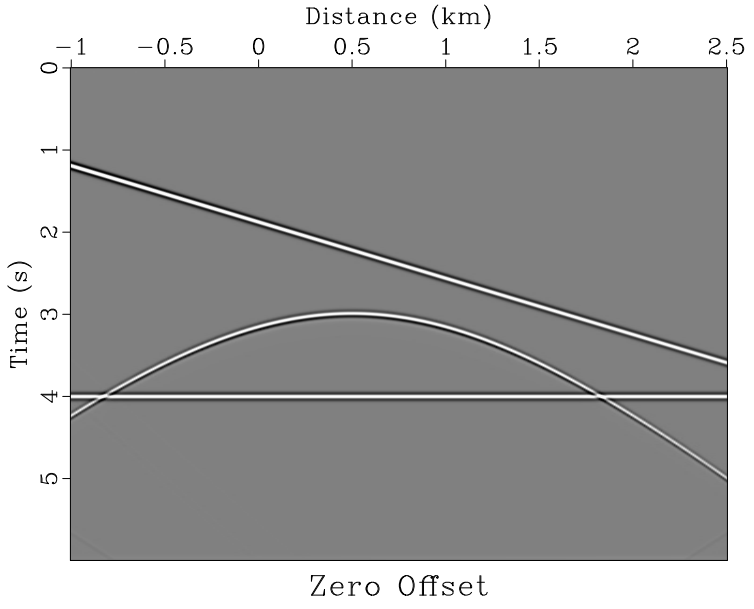
The slope decomposed initial migration from Figure 3.2 is propagated through

a suite of plausible migration velocities. I illustrate the initial migration and example velocities of 0.75 km/s, 1.0 km/s (the correct velocity), and 1.25 km/s. Slope gathers showing this process are shown in Figure 3.3. Stacking these propagated images over slope produces the image in Figure 3.4.

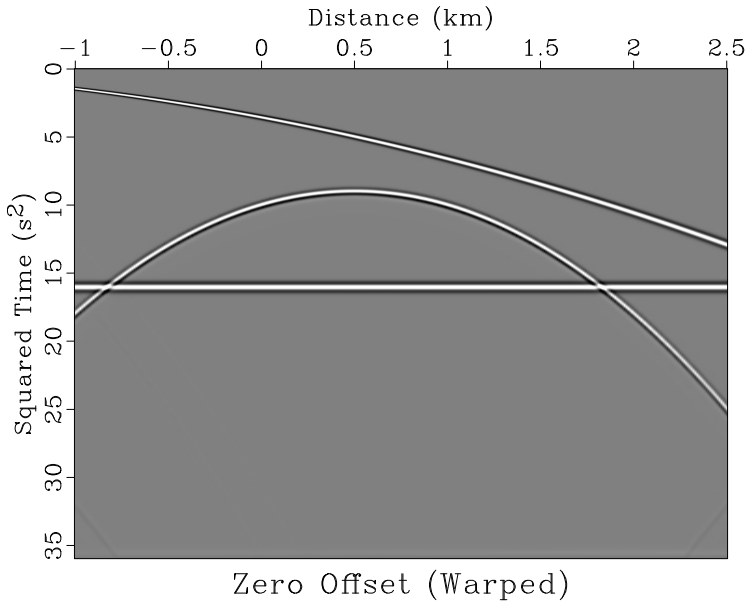
Examining the slope gather of the initial migration in Figure 3.3(a), the three panels contain points of energy corresponding, from top to bottom, to the top reflector, the diffractor, and the bottom reflector. The energy of each reflector is contained at the same lateral position in the three panels due to the constant slope of the reflector, although the vertical position of the top dipping reflector changes through the panels because the reflector dips downward to the right.

The diffraction has a hyperbolic moveout rather than a constant slope, so energy appears at different slopes in different slope gathers, with zero slope in the gather centered over the diffractor at 0.5 km. As I propagate data through velocity, this pattern holds: the reflection energy is stationary at its slope location for all gathers and diffraction energy has zero slope when viewed in the gather above the diffractor and non-zero slope for other gathers.

The initially migrated image is propagated to the higher time migration velocity of 0.75 km/s using oriented velocity continuation, and slope gathers are illustrated in Figure 3.3(b). Reflection energy now bends upward about the stationary point of each reflection in "smiles" that become more accentuated with larger migration velocities. The diffraction event bends upward as well, but this only holds for the current case of under-migration.



(a)



(b)

Figure 3.1: Toy model data: (a) zero-offset synthetic data featuring two sloping reflectors and a diffractor centered at 0.5 km; (b) zero-offset data warped to squared time.

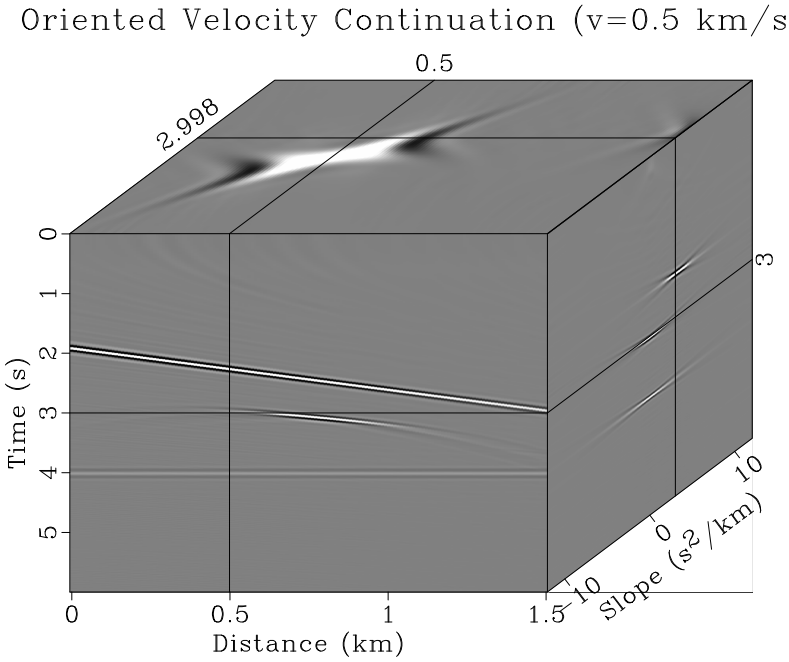


Figure 3.2: Slope decomposed initial migration for toy model data using $v_0 = 0.5$ km/s. The side pane shows the slope decomposed data centered at 0.5 km, the diffractor, and the front pane shows a partial image for data containing energy with a slope of 2.998 km/s^2, that of the top reflector.

Figure 3.3(c) shows the image propagated to the correct migration velocity. Diffraction energy is planar in all three gathers and flat in the middle gather centered above the diffractor. Stacking the energy in each gather over slope collapses the flat diffraction energy in the central gather to a point at the location of the diffractor. Sloping diffraction energy in the right and left panels cancels out when summed over slope. This flatness is essential to using oriented velocity continuation as a tool for determining the correct migration velocity.

When data are further propagated to over-migration with velocity of 1.25 km/s in Figure 3.3(d), the diffraction event bows downward in a "frown" juxtaposed against the upward bending reflection "smiles".

Stacking over slope provides images for these four velocities (Figure 3.4). In these images, the diffraction event incrementally evolves from having a hyperbolic downward character in the top under-migrated row, to collapsing to a point in the bottom left panel with the correct migration velocity, to bowing hyperbolically upward in the bottom right over-migrated image.

The changing geometry of diffraction energy in the gathers can be harnessed to determine the proper migration velocity (Landa et al., 2008; Reshef and Landa, 2009). The correct migration velocity will be the one that maximizes the "flatness" of slope decomposed diffraction events, as measured by coherence or another appropriate metric. Selecting 1.0 km/s, the velocity that produces flat diffraction energy, provides a properly migrated image (Figure 3.5).

To properly estimate migration velocity using diffraction flatness, reflection

events must first be filtered out from the diffraction data, or else the contribution of reflection "smiles" may dominate and bias the flatness measure.

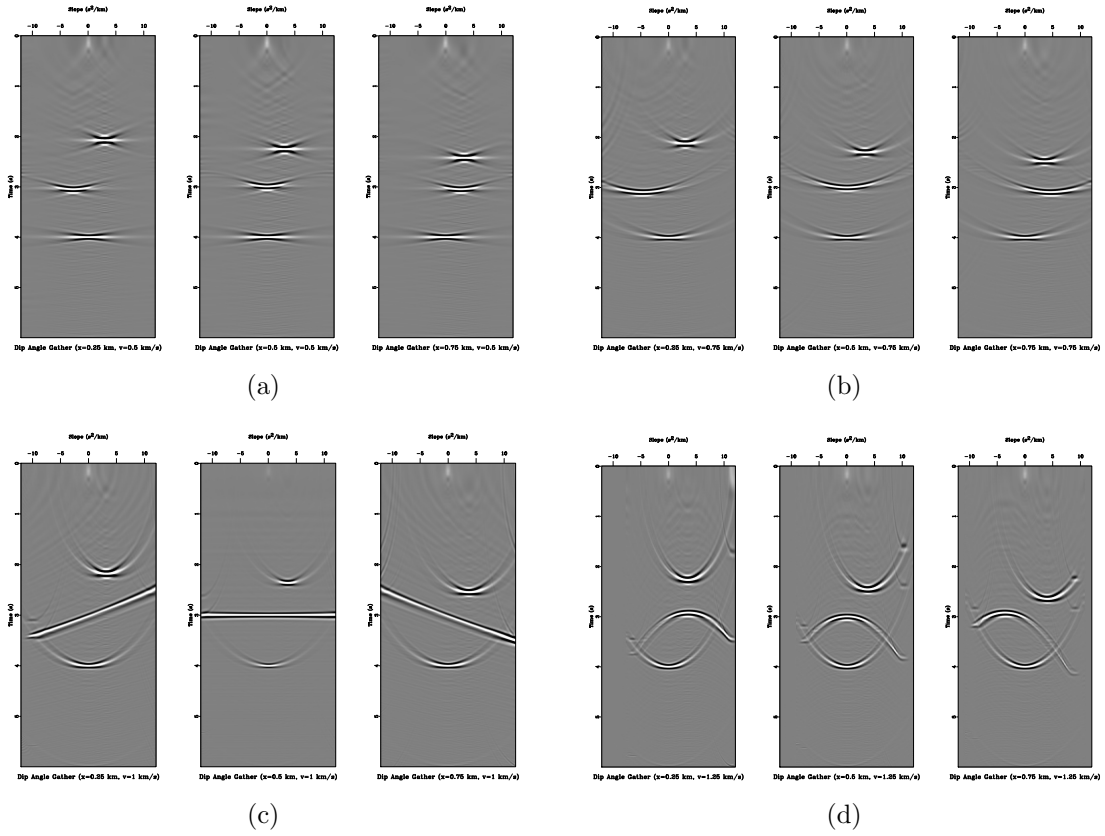


Figure 3.3: Slope gathers centered at 0.25 km (left), 0.5 km (center), and 0.75 km (right) for migration velocities of: (a) initial under-migration with 0.5 km/s; (b) under-migration with 0.75 km/s; (c) the correct migration velocity of 1.0 km/s; (d) over-migration with 1.25 km/s.

Synthetic Example

To test oriented velocity continuation and its velocity resolution I generate a synthetic dataset using a model with a constant velocity gradient beginning with

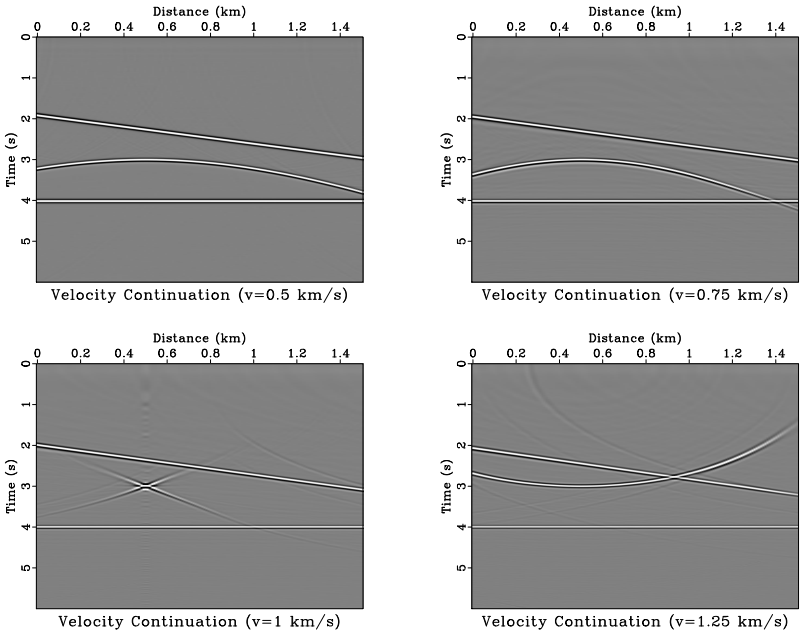


Figure 3.4: Toy model data propagated through oriented velocity continuation for different migration velocities.

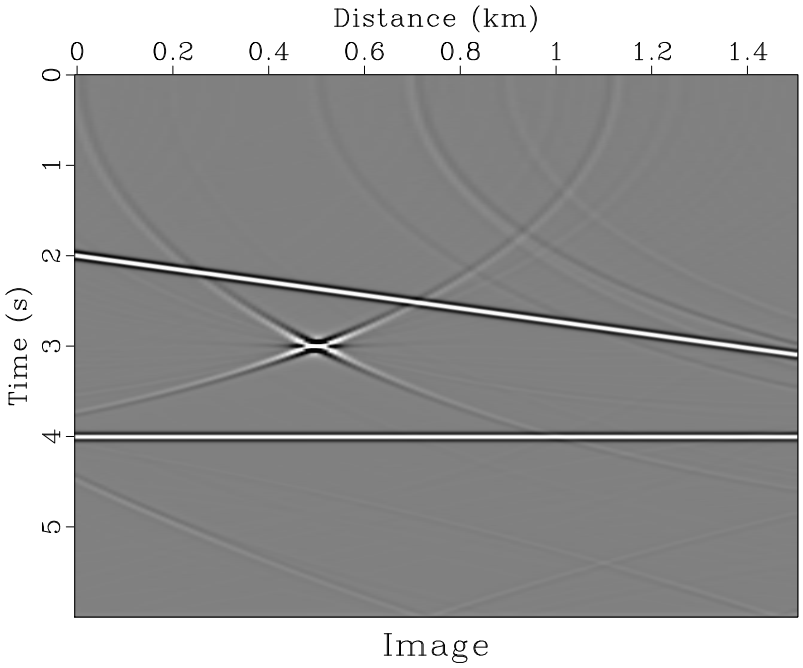
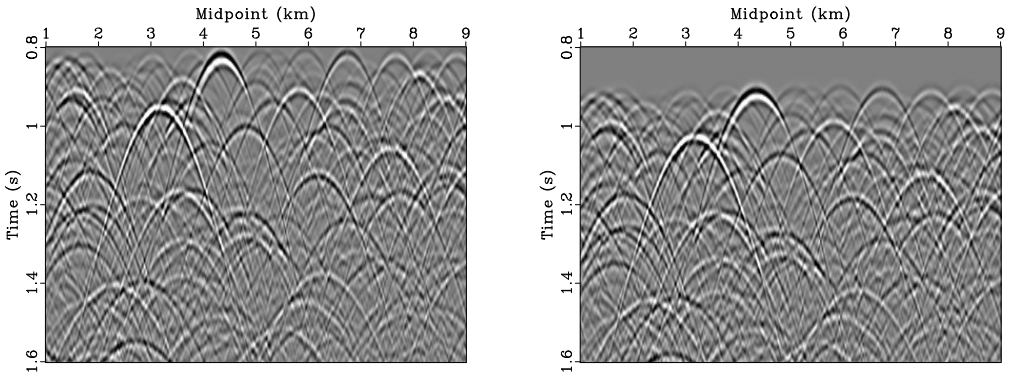


Figure 3.5: Toy model image using 1.0 km/s migration velocity.

a 2.0 km/s surface velocity. Diffractors are created as reflectivity spikes within the model with random spatial and magnitude distributions. Kirchhoff forward modeling is used to generate 24 offsets with a 50 m interval.



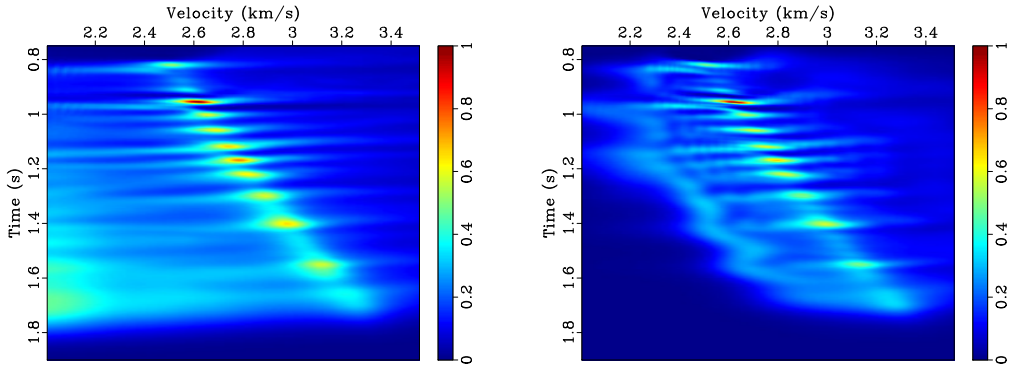
(a) (b)

Figure 3.6: Modeled data sections: (a) zero-offset; (b) 1.0 km offset section.

Zero-offset data are shown in Figure 3.6(a), and 1.0 km common-offset data appear in Figure 3.6(b). A time shift between the data is noticeable.

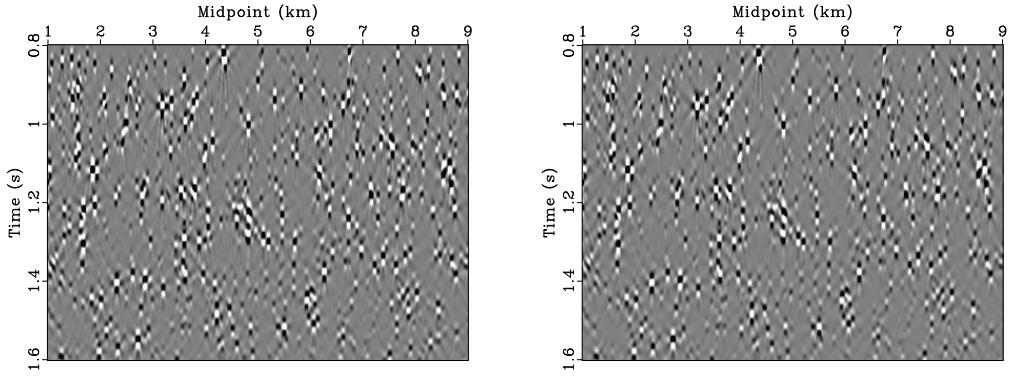
Both zero and common-offset data are warped to squared time, slope decomposed, and migrated with a 2.0 km/s initial velocity. The initially migrated slope decomposed images are propagated through a range of plausible migration velocities using oriented velocity continuation. Common-offset partial images are then stacked over offset for each continuation velocity.

Figure 3.9 illustrates slope gathers for zero and 1.15 km offsets generated for the image location $x = 2.32$ km with migration velocities 2.1, 3.0, and 3.9 km/s. 3.0 km/s is the correct velocity for the diffractor at 1.4 s located directly underneath the midpoint of this gather. When a different velocity is used, the shape of the



(a) (b)

Figure 3.7: Velocity scan semblance panels calculated for: (a) zero-offset; (b) all 24 offsets (50 m interval).



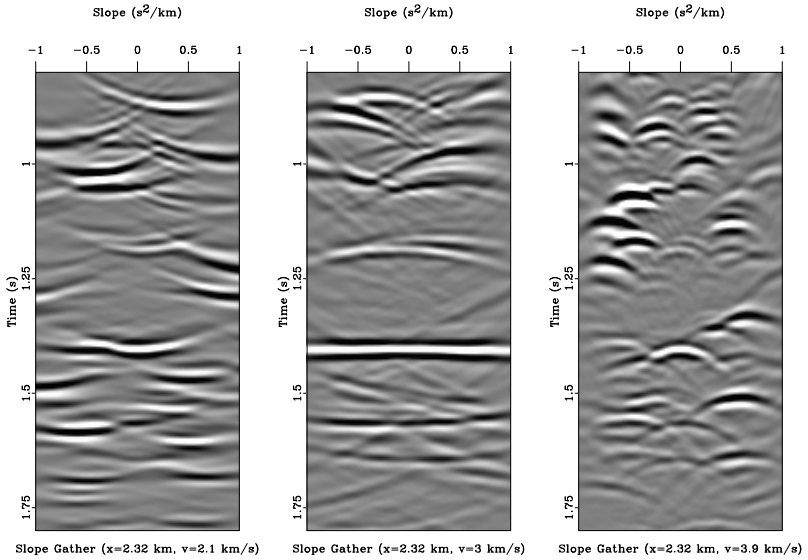
(a) (b)

Figure 3.8: Images: (a) zero-offset OVC (velocity from Figure 3.7(a)); (b) stacked common-offset OVC (velocity from Figure 3.7(b)).

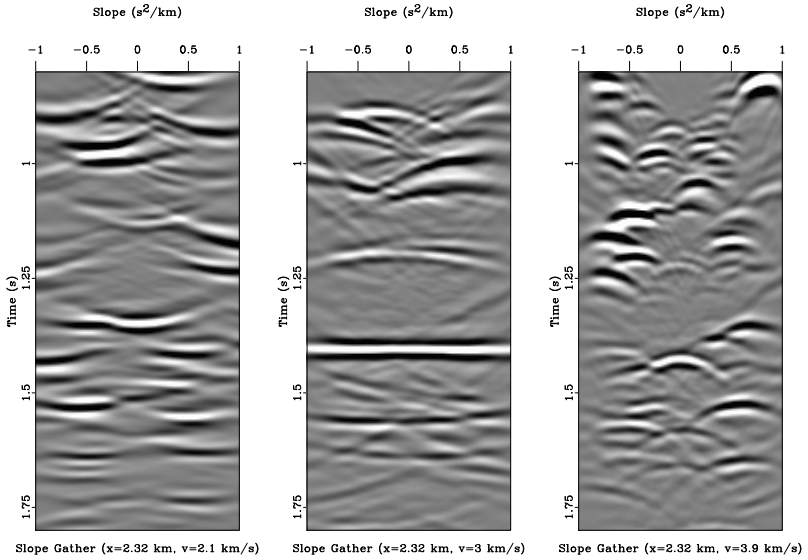
event deviates from planar. I perform velocity analysis by testing the semblance, or flatness, of diffraction events in slope gathers over the range of velocities. Because velocity does not vary laterally in this synthetic model, I average the semblance across midpoints to generate semblance panels for the zero-offset and common-offset cases (Figures 3.7(a) and 3.7(b) respectively).

As seen from the slope gathers (Figure 3.9), for the zero-offset case, there is a stationary point corresponding to diffraction energy with zero slope which does not shift vertically under velocity perturbations. For the 1.15 km common-offset case, perturbing velocity changes the slope decomposed diffraction shape and shifts it vertically. Slope gathers with incorrect velocities (Figure 3.9(b)) are time shifted with respect to those generated for the zero-offset case (Figure 3.9(a)). When the correct migration velocity is used, horizontal common-offset diffraction energy appears at the same time as for the zero-offset case. The vertical shift of incorrectly migrated common-offset data leads to a sharper change in estimated flatness values while converging on the correct migration velocity and therefore improves velocity resolution. Therefore, common-offset semblance panel appears to have higher spatial and vertical resolution than the zero-offset case. This higher spatial resolution can be attributed to the improved illumination of scattering objects with the full range of offsets.

Final images for zero-offset and stacked common-offset cases using migration velocities estimated from the semblance panels are shown in Figure 3.8. Differences between the two images are too small to easily detect in this example. However, due to the higher velocity resolution visible in the semblance panel resulting from the consideration of multiple offsets, Figure 3.7(b), I expect the stacked common-offset



(a)



(b)

Figure 3.9: Slope gathers: (a) zero-offset; (b) 1.15 km offset.

image to be better resolved and less prone to noise than zero-offset case when applied to field datasets.

Field Data Example

I demonstrate an application of zero-offset (post-stack) oriented velocity continuation on a deep water 2D line acquired to image the Nankai Trough subduction zone. Data acquisition parameters as well as processing results can be found in Moore et al. (1990), where the line is referred to as NT62-8. Structural interpretation can be found in Moore and Shipley (1993). Here, I consider a fragment of the line (CMPs 900-1301) used previously by Forel et al. (2005).

Conventional velocity analysis resolution suffers in this dataset from the limitations imposed by the depth of a seabed in the area (average of $\approx$ 4.5 km) and a relatively short 2 km streamer length. For deep water datasets diffractions may exhibit better illumination than reflections because diffraction aperture is not restricted to the recording array length, enabling them to provide a potentially more detailed velocity distribution. This behavior makes OVC migration velocity analysis appealing.

The DMO stacked section considered in this study is shown in Figure 3.10. Diffractions are extracted via plane-wave destruction (Figure 3.11), warped to squared time, and decomposed into slope. Figure 3.12 shows slope decomposed data warped back to regular time for ease of comparison with slope decomposed images appearing later. Next, I take the decomposed data through oriented velocity continuation over

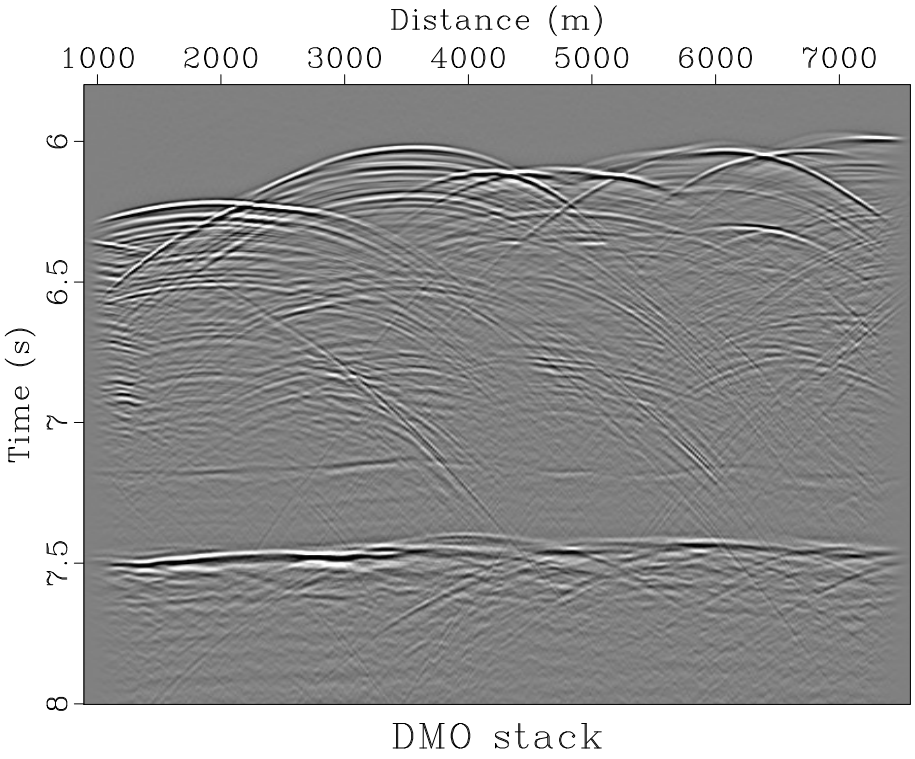


Figure 3.10: Nankai DMO stacked section.

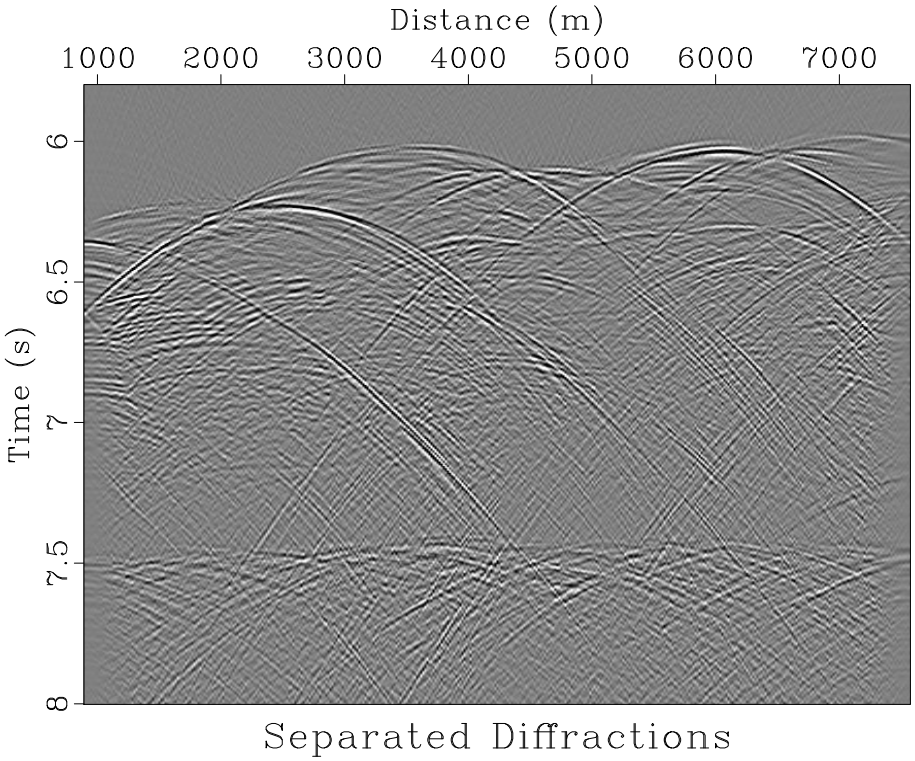


Figure 3.11: Nankai separated diffractions.

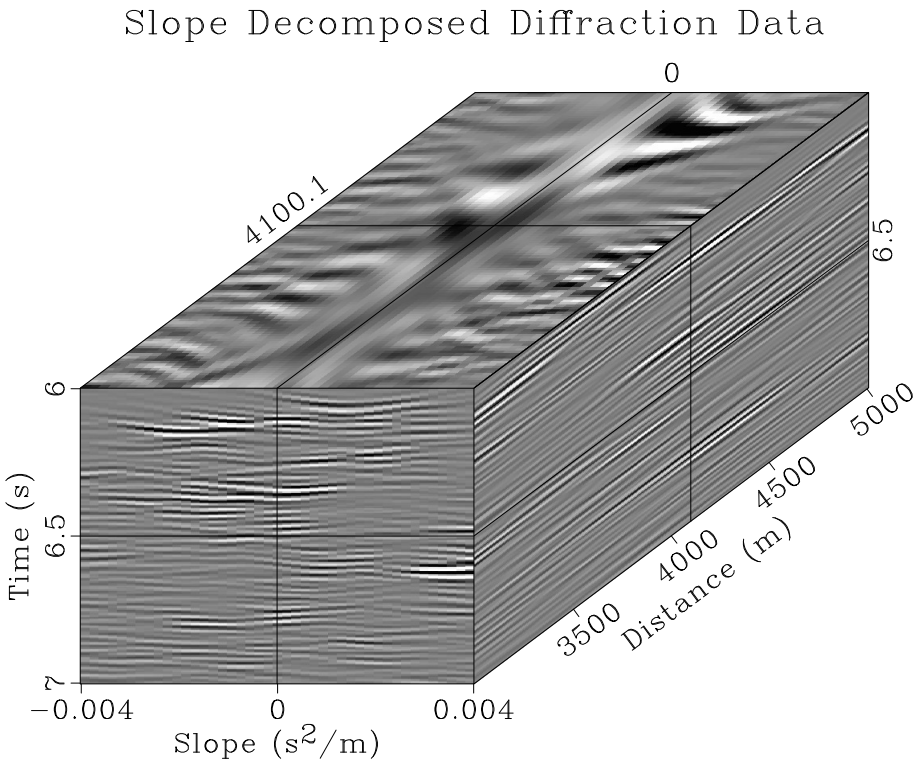


Figure 3.12: Slope decomposition of Nankai diffraction data.

a range of sixty constant migration velocities beginning with $v_0 = 1.4$ km/s using a 20 m/s step. Diffraction events bend upward in the slope gather centered above $x = 4100.06$ m with the minimum tested migration velocity (Figure 3.14(a)), indicating under-migration. Diffraction events in the slope gather centered above the same location with the maximum tested migration velocity (Figure 3.14(b)) bend downward, indicating over-migration.

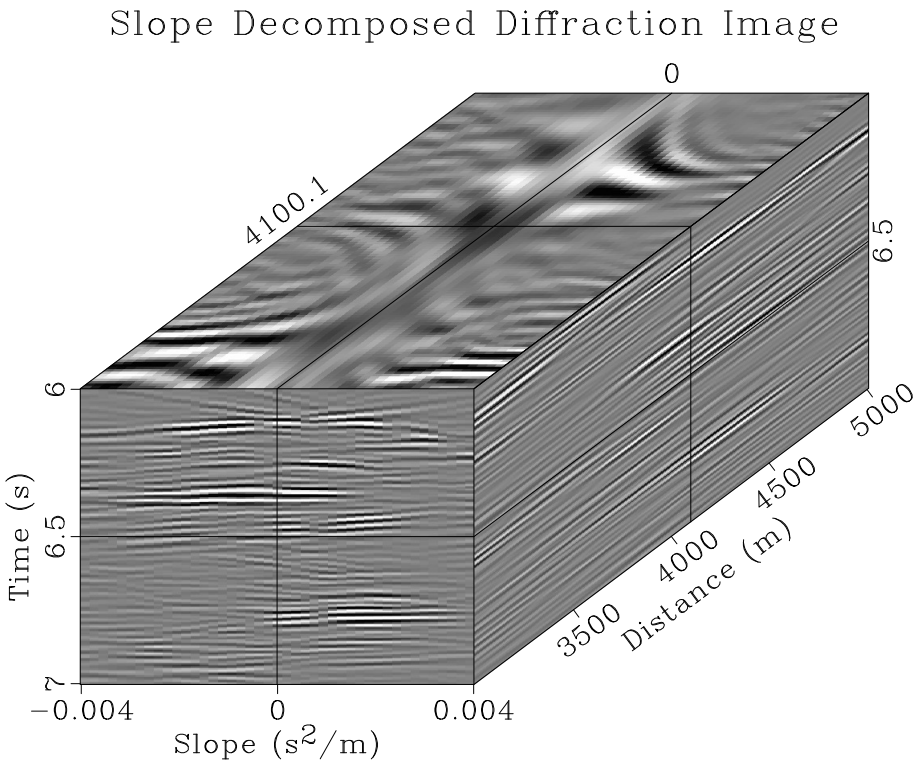


Figure 3.13: Slope decomposition of Nankai diffraction image

Gather semblance is calculated for each continuation velocity, and migration velocity is automatically picked by attempting to maximize semblance for plausible velocity values at each CMP location. Semblance panels with superimposed picks are

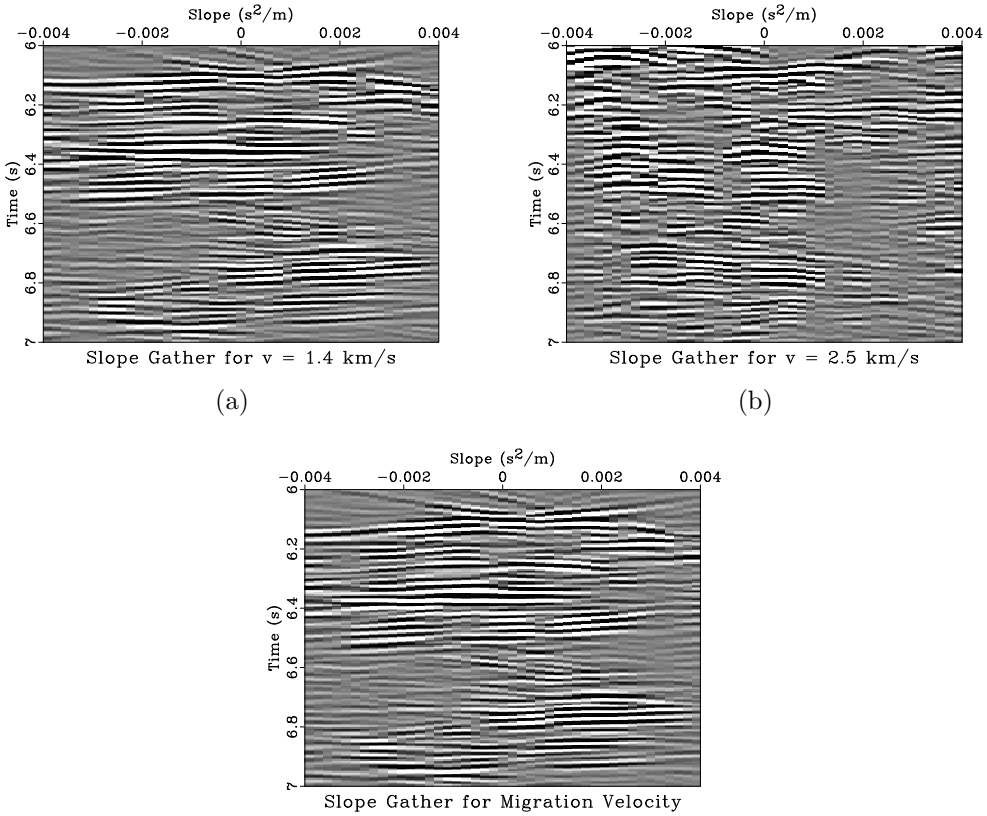


Figure 3.14: Slope gathers centered above $x = 4100.06$ m migrated with: (a) 1.4 km/s, (b) 2.5 km/s and (c) picked migration velocity.

shown in Figure 3.15. Anomalies corresponding to higher velocities than the picked trend may correspond to reflections with high curvature, like the one located between $x = 3000$ and 4000 m between $t = 6.0$ and 6.5 s in Figures 3.10 and 3.11. Highly curved reflections have a similar behavior to diffractions in response to migration velocity perturbation (Sava et al., 2005), but focus at a higher velocity than the correct one. Intersection of over-migrated reflection and diffraction tails from the rugose seabed, some of which are out of plane, leads to diffraction-like events, another cause of false semblance highs. These are visible in the three semblance panels of Figure 3.15 above 6.4 s. Low velocity semblance anomalies corresponding to the flattening of out of plane diffractions are also visible in the middle interval of the semblance panels, particularly near $t \approx 6.8$ s in the central and right panels centered above $x = 5000$ and 6500 m.

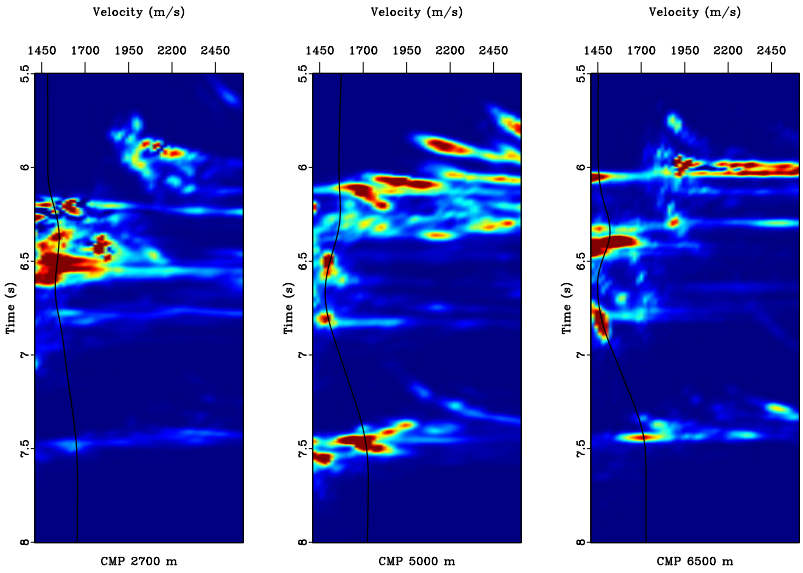


Figure 3.15: Velocity scan semblance panels with superimposed picks from left to right for CMPs at 2700 m, 5000 m, and 6500 m.

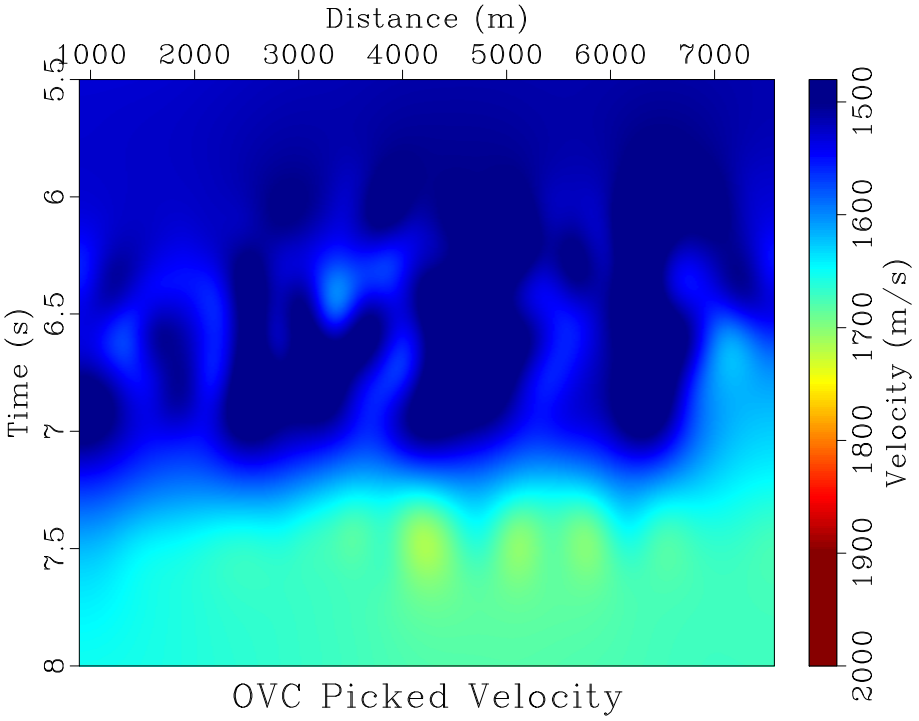


Figure 3.16: Velocity picked from slope-gather flattening.

Combining the semblance velocity picks from each CMP provides a time-migration velocity field, shown in Figure 3.16. As noted above, several anomalously low velocity zones exist in the picked field, primarily between $t = 6.5$ and 7 s where the attempted flattening of out of plane diffractions leads to a low picked velocity.

Gathers corresponding to the picked velocity are selected. Examining a slope gather from $x = 4100.06$ m generated using the picked migration velocity (Figure 3.14(c)), diffraction events now appear flat, particularly the one located near $t = 6.4$ s, indicating that they have been correctly migrated.

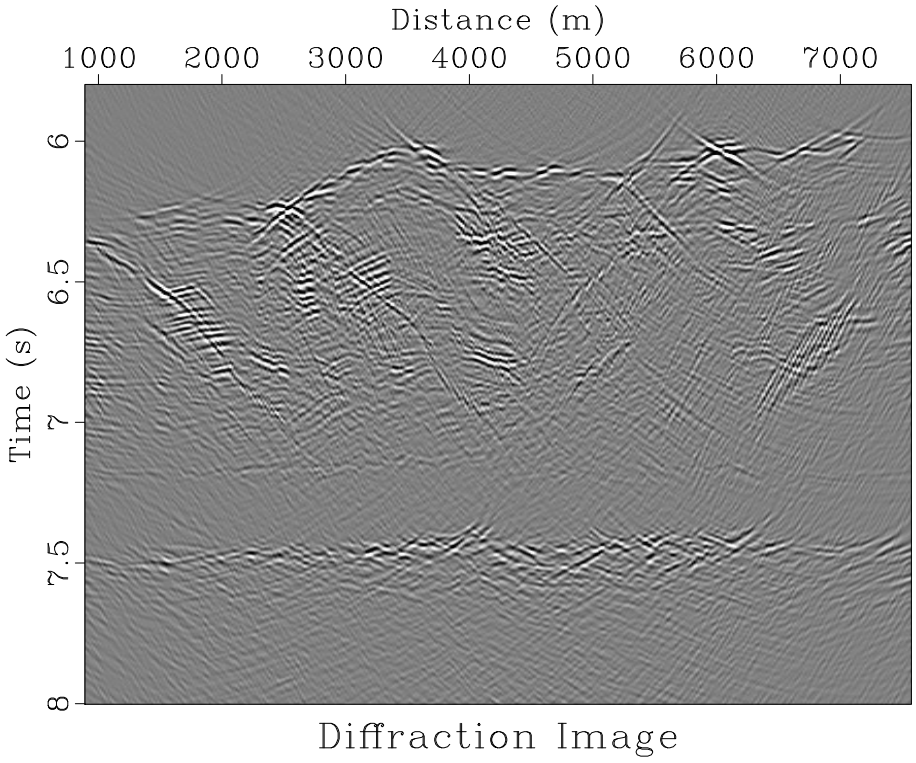


Figure 3.17: Diffraction image generated with the velocity from the Figure 3.16.

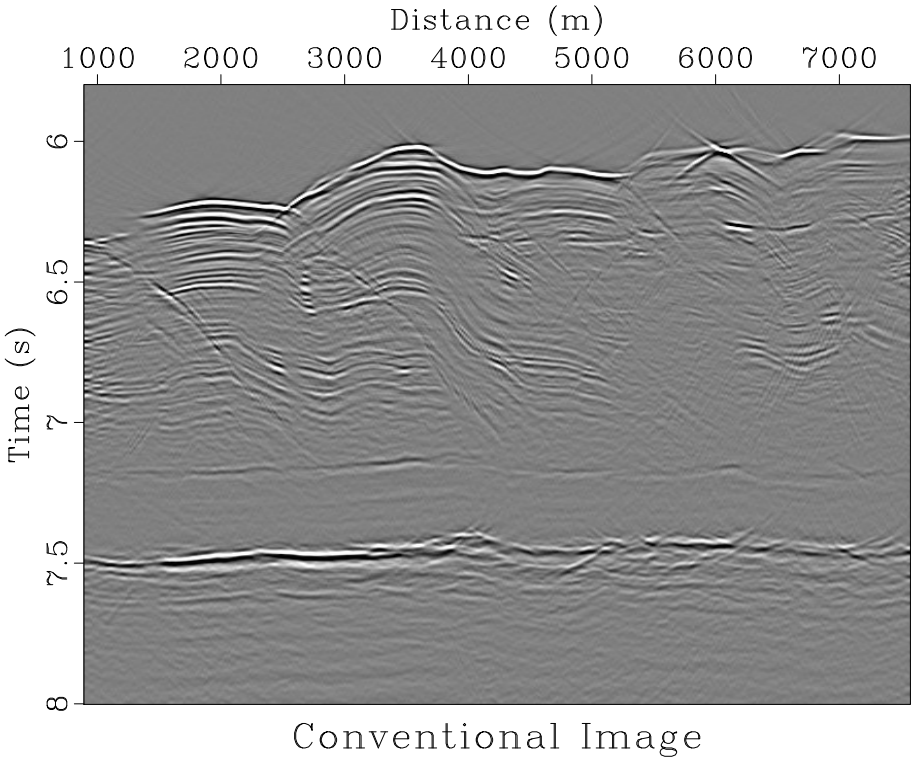


Figure 3.18: Conventional image generated with the velocity from the Figure 3.16.

Stacking gathers generated from the the picked velocity over slope provides the diffraction image in Figure 3.17. I apply oriented velocity continuation to the DMO stacked data from Figure 3.10 and stack over gathers selected with the appropriate velocity to generate the image of reflections and diffractions in Figure 3.18. Both images highlight fault surfaces. Finer discontinuities, such as those associated with the rough surface of the subducting plate crust, located near $t \approx 7.5$ s (Moore and Shipley, 1993), are more prominent on the diffraction image and tend to be well focused, supporting the accuracy of the picked velocity.

CONCLUSIONS

I have developed and demonstrated a highly parallel and constructive procedure for time-domain velocity estimation. The method operates by decomposing data by slope and propagating slope components in velocity in the midpoint-time-slope domain. Semblance in slope gathers is used as a measure for selecting velocities that correspond to correctly migrated flat diffraction events, which applies even for single-offset data. This semblance measure is observed to achieve higher resolution when multiple offsets are considered in the oriented velocity continuation process. If data with multiple offsets are available, oriented velocity continuation could be used to better constrain the velocity model when used in conjunction with traditional reflection moveout analysis. Chosen velocities can be used to generate both diffraction and reflection images. The powerful ability of oriented velocity continuation to operate with only zero-offset data enables accurate migration velocity analysis in situations where only limited-offset data are available.

Flatness of slope-decomposed diffraction events is more responsive to velocity perturbation than diffraction focusing, because it does not require smoothing or windowing in space. Therefore, the proposed method has the potential for diffraction velocity estimation with superior resolution when compared to methods based on diffraction focusing.

Oriented Velocity Continuation is formulated as a type of time-migration, and is thus subject to constraints relating to image distortion from horizontal velocity changes in the subsurface. The presence of strong lateral velocity variations may alter diffraction moveout, making event slope change with azimuth. In such a case, OVC would be unable to completely flatten the diffraction signal in slope gathers and locally determine the correct migration velocity.

This method can be extended to three dimensions using data decomposition by azimuth and inclination for each image point. Operating on three dimensional data should improve velocity resolution by overcoming out of plane artifacts in the seismic image. Although extension to 3D adds the expense of additional spatial and slope dimensions, the Fourier-domain computation would also be parallel in these new dimensions, making the operation feasible in practice using computer clusters.

Chapter 4

A probabilistic approach to seismic diffraction imaging

I propose and demonstrate a probabilistic method for imaging seismic diffractions based on path-integral imaging. Our approach utilizes oriented velocity continuation to produce a set of slope-decomposed diffraction images over a range of plausible migration velocities. Utilizing the assumption that each partial image in slope is independant enables us to construct an object resembling a probability field from the slope-decomposed images. That field may be used to create weights for each partial image in velocity corresponding to the likelihood of a correctly migrated diffraction occurring at a location within the seismic image for that migration velocity. Stacking these weighted partial images over velocity provides us with a path-integral seismic diffraction image created using probability weights. I illustrate the principles

[1]Some of the material in this chapter was published or submitted as:

- Decker, L., and S. Fomel, 2019, Path-integral seismic diffraction imaging with probability weights: SEG International Exposition and Annual Meeting, 4231–4235. The authors contributed to this paper as follows: study conception and design: L. Decker; data collection: L. Decker; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.
- Decker, L., and S. Fomel, 2021b, A probabilistic approach to seismic diffraction imaging: Lithosphere, In Review. The authors contributed to this paper as follows: study conception and design: L. Decker; data collection: L. Decker; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.

of the method on a simple toy model, show its robustness to noise on a synthetic, and apply it to a 2D field dataset from the Nankai Trough. I find that using the proposed approach creates diffraction images that enhance diffraction signal while suppressing noise, migration artifacts, remnant reflections, and other portions of the wavefield not corresponding to seismic diffraction relative to previously developed diffraction imaging methods, while simultaneously outputting the most likely migration velocity. The approach outlined in this paper is complimentary to existing data domain methods for diffraction extraction, and the probabilistic diffraction images it outputs can supplement existing reflection and diffraction imaging methods by highlighting features that have a high likelihood of being diffractions and accentuating the geologically interesting objects in the subsurface that cause those features.

INTRODUCTION

Though less popular than reflection imaging for characterizing the subsurface, diffraction imaging has been gaining increasing attention (Landa, 2012). Seismic diffractions occur when a seismic wave interacts with an object on the order of its wavelength, such as a fault, fracture, or void (Harlan et al., 1984; Fomel et al., 2007; Moser and Howard, 2008; Klokov and Fomel, 2012; Decker et al., 2015; Popovici et al., 2015; Schwarz, 2019), and may even be able to resolve objects beyond the seismic wavelength (Khaidukov et al., 2004).

Seismic diffractions are significantly weaker than reflections (Klem-Musatov, 1994), requiring their quarantining from the stronger reflection signal to be usable. Numerous methods exist for separating diffraction signal from reflection, in both

the data and image domains (Harlan et al., 1984; Kozlov et al., 2004; Fomel et al., 2007; Moser and Howard, 2008; Berkovitch et al., 2009). Attributes correlated to a diffraction occurring, like focusing (Khaidukov et al., 2004) or angle-gather flatness (Landa et al., 2008; Reshef and Landa, 2009) only become available after migration, so these methods typically function by predicting and removing reflections, although recent excitement in applying machine learning techniques to problems related to seismic imaging and interpretation (Pham et al., 2019; Kaur et al., 2019; Wu et al., 2020) has extended into the use of pattern recognition (de Figueiredo et al., 2013 and deep learning (Tschannen et al., 2020) for diffraction detection. Reflection removal leaves the diffractions but also the already present noise. Diffraction images often feature poor signal to noise ratios and situations where it is difficult to distinguish what features are diffraction, noise, or migration artifacts (Harlan et al., 1984; Fomel et al., 2007; Decker et al., 2013). This difficulty motivated Decker et al. (2017b) to treat the semblance of diffraction image angle gathers as a proxy for diffraction likelihood at different locations in space, and use that semblance as a model weight for least-squares Kirchhoff migration (Nemeth et al., 1999). However, in practice this involved thresholding semblance at a selected level, making the process both arbitrary and non-linear. To overcome this issue I turn to path-integral seismic imaging.

Path-integral seismic imaging (Landa et al., 2006) applies the stationarity of Feynman path integrals to the problem of seismic time imaging to create images without knowing a velocity model. Integrating (stacking) over all possible seismic signal paths, which are dependent on velocity, with an appropriate weighting function produces a seismic image based on the concept that the seismic image is stationary

at the correct velocity. Thus, paths with incorrect velocity interfere destructively during summation whereas correct velocity signal interferes constructively. Schleicher and Costa (2009) utilized this concept to determine seismic migration velocity, while Burnett and Fomel (2011), Merzlikin and Fomel (2017), Merzlikin et al. (2019), and Merzlikin et al. (2020) applied the technique to seismic diffraction imaging.

I observe that if the weight functions used in the path-integral imaging equations are treated as probability distributions, the imaging and velocity analysis techniques become equivalent to calculating expectation values for the time image and time migration velocity respectively (Fomel and Landa, 2014b). This is immediately clear when examining the imaging condition of Landa et al. (2006), which treats the imaging process as a weighted summation of images within a set corresponding to the probability of each. This observation inspires us to use attributes corresponding to diffraction probability as weight functions. In order to utilize such a framework I first need a suite of seismic images over different migration velocities and their angle gathers, which are provided by oriented velocity continuation (OVC). OVC involves the continuous propagation of slope-decomposed seismic images along their characteristics over different time migration velocities (Decker and Fomel, 2014; Decker et al., 2017a). Other continuation operators exist, for example Burnett et al. (2011) proposed a method for applying azimuthally anisotropic velocity continuation to zero-offset data. The slope-decomposed partial images produced by OVC are equivalent to dip-angle gathers (Decker and Klokov, 2014), enabling us to compute gather semblance and other path-integration weights. OVC provides us with an ensemble of slope-decomposed seismic images over a range of different time migration velocities.

I propose to use these images to construct the weight functions for path integration.

This paper demonstrates the principles of our probabilistic imaging process and applications to demonstrate its utility. In the next section I outline how I may construct weight functions for path-integral imaging correlated to the probability of a correctly migrated diffraction occurring at a location in space for a given velocity. I illustrate the methodology on a toy model example and then apply it to a synthetic data set to illustrate its robustness to noise relative to deterministic methods for diffraction imaging and equal weight path-integral imaging. The method is applied to a field data set of the Nankai trough to show how it is able to both generate a diffraction image with suppressed noise and output a plausible velocity field that is able to highlight geologically interesting features like a velocity inversion. Finally, I end the paper with conclusions and outlining some promising avenues for future work.

THEORY

The imaging condition of Landa et al. (2006) for weighted path-integral image Q_w, weights $w(\eta)$, stack $Q(\eta)$ and wavefront multiparameter η, which parameterizes all possible ray paths may be written as:

$$Q_w = \int w(\eta)\ Q(\eta)\ d\eta. \tag{4.1}$$

Allowing $w(\eta)$ to be a probability distribution, Q_w becomes an image expectation value. For the case of time migration, ray paths are uniquely parameterized by veloc-

ity, so η in the above equation may be replaced by v. A deterministic image would use a Dirac delta at the most likely velocity as the weight function. To create our weight functions I will apply Oriented Velocity Continuation (OVC) to generate a collection of slope-decomposed seismic images over different migration velocities, $I(\mathbf{x}, v, p)$, where $\mathbf{x}$ is a vector describing the position within a seismic image, v is the migration velocity, and p is the image slope. If I consider each slope-decomposed seismic trace within a gather at position $\tilde{\mathbf{x}}$ at velocity $\tilde{v}$, $I(\tilde{\mathbf{x}}, \tilde{v}, p)$, to be independent, I may calculate gather semblance and treat it as a likelihood of a properly migrated seismic diffraction event occurring at that location. This is based on the observation that diffractions migrated with correct velocity possess flat angle-gathers. The semblance, $\alpha \in [0, 1]$, which will serve as our first weight for path integration, may be calculated as:

$$\alpha(\mathbf{x}, v) = \frac{\left(\int I(\mathbf{x}, v, p)\, dp\right)^2}{\int I(\mathbf{x}, v, p)^2\, dp}, \tag{4.2}$$

$$W_1(\mathbf{x}, v) = \alpha(\mathbf{x}, v). \tag{4.3}$$

Note that unbounded integrals are considered to be over the whole domain of the integrating variable. To avoid issues related to dividing by 0, all division operations are treated as an inversion involving shaping regularization (Fomel, 2007a). Semblance may be used additionally to calculate the expectation velocity $\bar{v}(\mathbf{x})$ at each location $\mathbf{x}$ and the corresponding velocity variance, $\sigma_v^2(\mathbf{x})$:

$$\bar{v}(\mathbf{x}) = \frac{\int v\, \alpha(\mathbf{x}, v)\, dv}{\int \alpha(\mathbf{x}, v)\, dv}, \tag{4.4}$$

$$\sigma_v^2(\mathbf{x}) = \frac{\int \left(\bar{v}\left(\mathbf{x}\right) - v\right)^2 \, \alpha(\mathbf{x}, v)\, dv}{\int \alpha(\mathbf{x}, v)\, dv}. \tag{4.5}$$

The denominator ensures proper normalization of α, whose integral over v is not necessarily equal to 1. Assuming normally distributed diffraction information around the expectation velocity allows us to construct our second weight, $W_2(\mathbf{x}, v)$:

$$W_2(\mathbf{x}, v) = \frac{\exp\left(-\frac{(\bar{v}(\mathbf{x})-v)^2}{2\sigma_v^2(\mathbf{x})}\right)}{\int \exp\left(-\frac{(\bar{v}(\mathbf{x})-v)^2}{2\sigma_v^2(\mathbf{x})}\right) dv}. \tag{4.6}$$

I construct the final weight, $W_3(\mathbf{x}, v)$ using the observation that properly migrated diffractions are focused, or localized in space, and therefore the magnitude of the spatial derivative of semblance normalized by the value of semblance at that location should change rapidly in the vicinity of properly migrated diffractions:

$$W_3(\mathbf{x}, v) = \frac{||\nabla_{\mathbf{x}} \alpha(\mathbf{x}, v)||_2}{\alpha(\mathbf{x}, v)}. \tag{4.7}$$

More weights may be used by eager practitioners able to determine other attributes correlated to diffraction likelihood, but seeking a simplification, I content myself with three. Thus, for m weights enumerated by j, and allowing $\tilde{I}(\mathbf{x}, v) = \int I(\mathbf{x}, v, p)\, dp$, I may calculate our probabilistic diffraction image as:

$$\widehat{I}(\mathbf{x}) = \int \tilde{I}(\mathbf{x}, v) \prod_{j=1}^{m} W_j(\mathbf{x}, v)\, dv. \tag{4.8}$$

Note that the weights I have constructed (and the combined weights) are pseudo-probabilities with the exception of W_2. They do not have all the properties of a true probability function, as that would imply that a diffraction must occur within a region in space. Rather, these weights show a general tendency for whether or not diffractions are likely to occur. Because W_2 is a probability of a velocity being accurate, and intuitively I assume that there must exist an accurate migration velocity at every point in space, I may build that weight so it fulfills the criteria of a probability.

METHODOLOGY

I illustrate our probabilistic diffraction imaging method on a toy model (Landa et al., 2008; Klokov and Fomel, 2013; Decker et al., 2017a). Figure 4.1 features a chart with the workflow for the imaging process. Note that when comparing this chart to the figures of the toy model, the vector $\mathbf{x}$ in the chart corresponds to $[t, x]$ in the toy model example. The order of arguments for objects in the chart and toy model may differ, as the order of arguments in figure labels correspond to the order axes are displayed in the figure. The toy model contains one reflector dipping downward to the right, a single point diffractor centered beneath 0.5 km and a planar reflector with a constant velocity of 1.0 km/s . Zero-offset data are modeled and displayed in Figure 4.2.

The process of OVC is applied to the data and their slope-decomposed images are propagated through a range of migration velocities to create a series of slope decomposed partial images. Slices through this volume are illustrated in Figure 4.3.

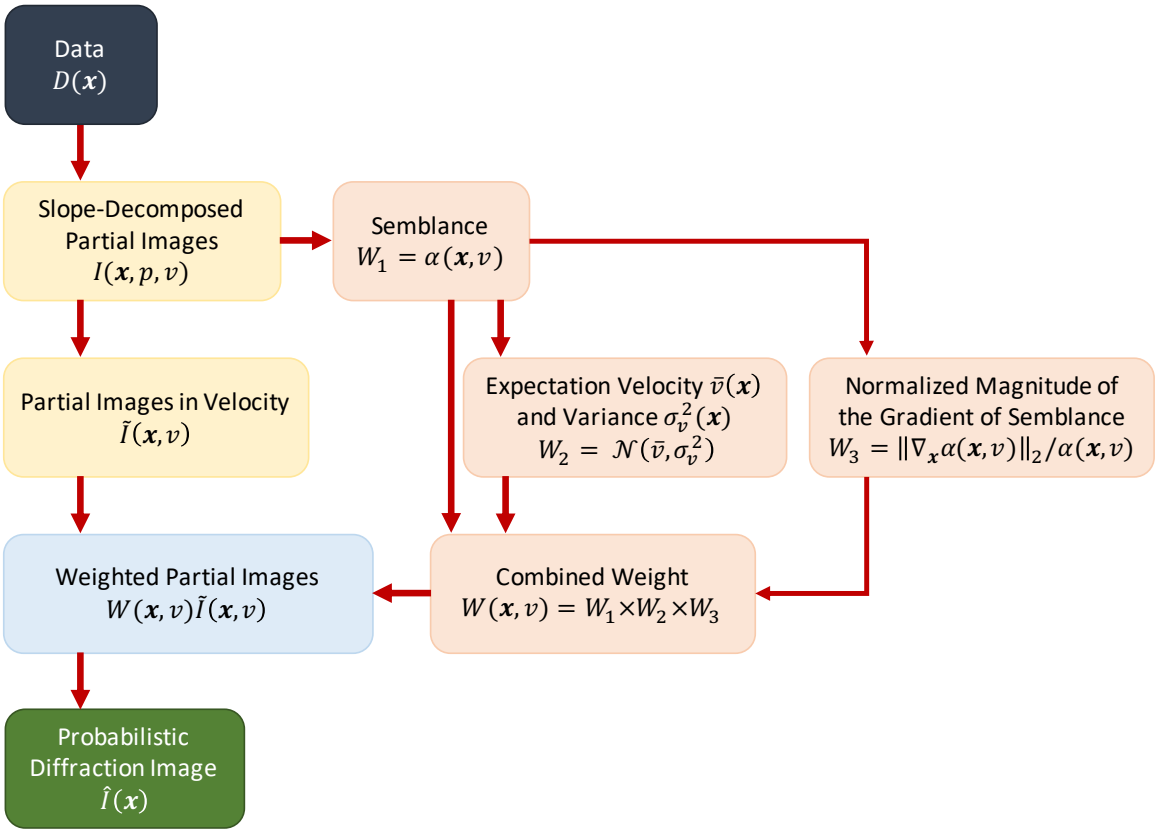


Figure 4.1: Probabilistic diffraction imaging workflow.

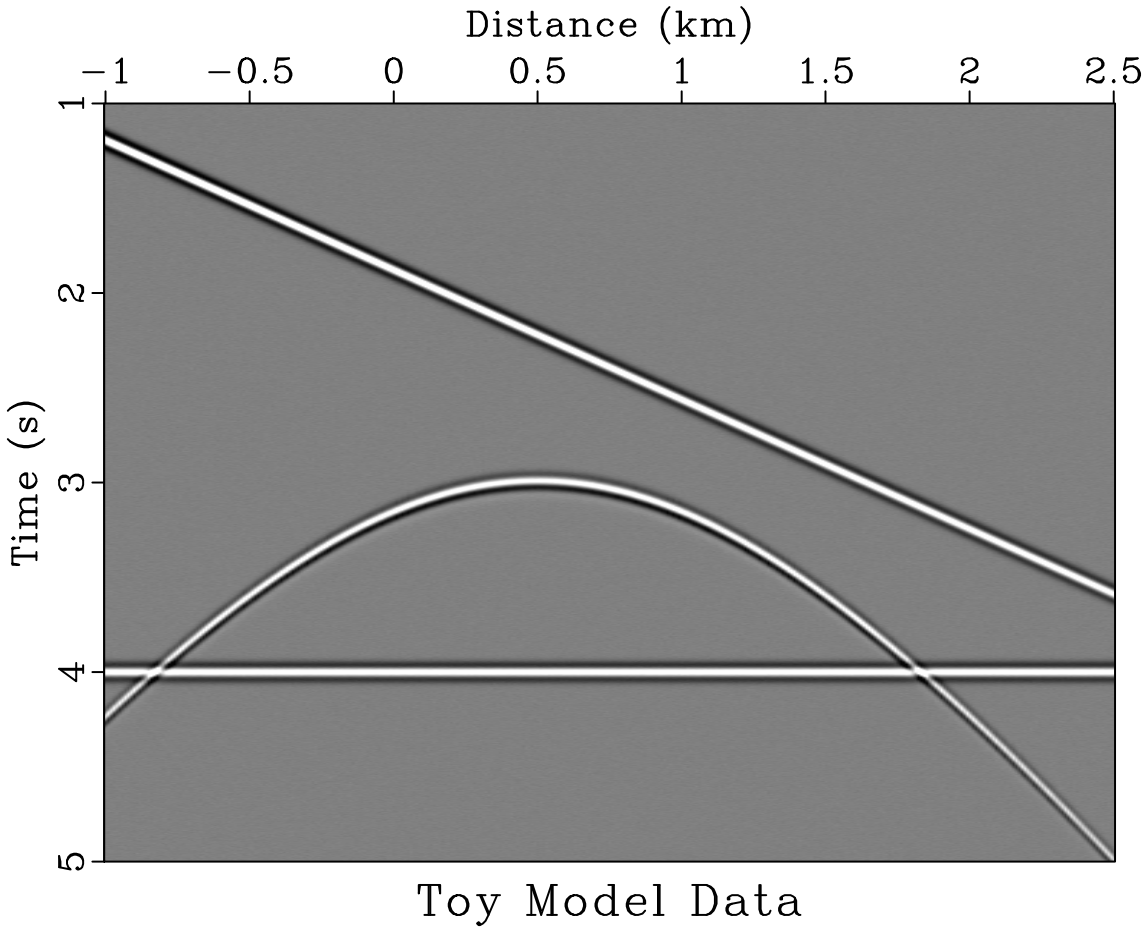


Figure 4.2: Toy model data containing two reflectors and a point diffractor.

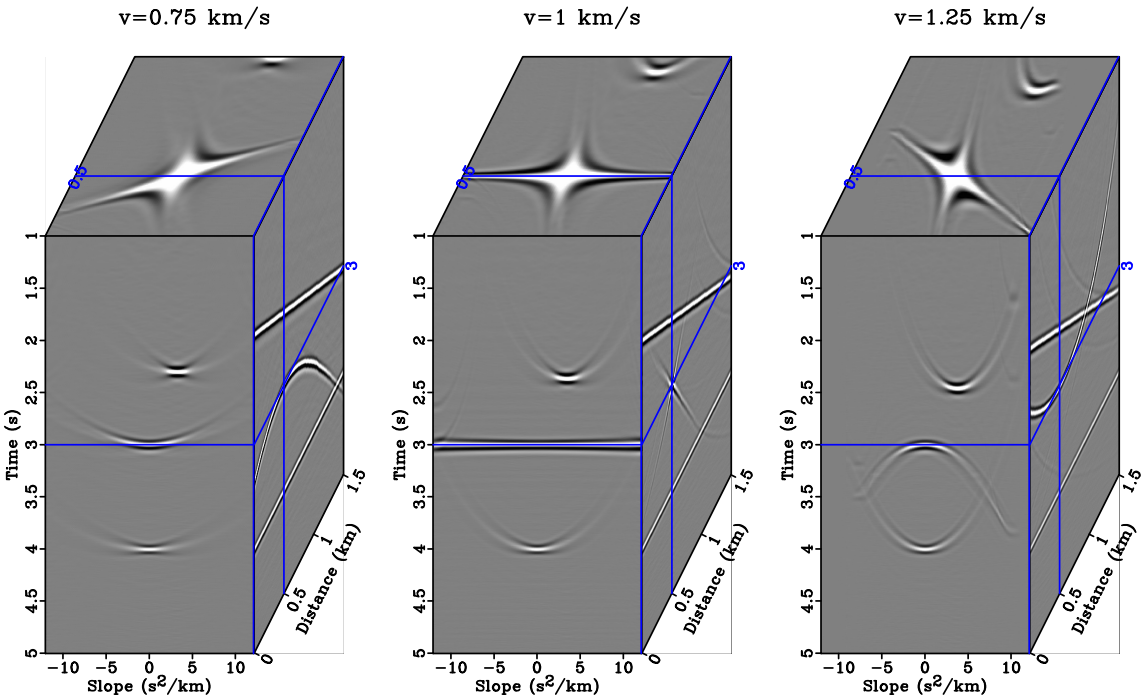


Figure 4.3: Example of application of OVC to toy model for three continuation velocities: left panel is under-migration with 0.75 km/s; central is for 1.0 km/s, the correct velocity; right is over-migration with 1.25 km/s. Front face of box plots show slope gathers centered above the diffraction. Right face of box plots show the image that results from stacking over slope gathers for that continuation velocity.

The front panes of box plots in this figure display slope gathers centered at 0.5 km, directly above the diffractor, for three different migration velocities. The right panes show the image created by stacking over slope for the selected migration velocity. Notice that for all three velocities, energy corresponding to the reflectors, near 2.1 s and 4 s, bend upward in the slope gathers. The lowest point in the upward bending reflection energy corresponds to the slope of that reflector. Notice that this apex is achieved at greater time values with higher migration velocities for the dipping reflector, and that the slope corresponding to this apex is greater for higher migration velocities. This is because migration makes dipping events steeper, and migration with larger velocities further increases the slope of dipping events. Diffraction energy in the left panel bends upward in a "smile" indicating under-migration. The energy flattens in the middle panel corresponding to correct migration. In the right panel it bends downward in a "frown" indicating over-migration. As I will see, the flat, correctly migrated diffraction energy results in a semblance high. Examining the stacked continuation images on the right panes of the box plots, I see that the diffraction bows downward when under-migrated in the left panel, is focused into a point when properly migrated in the center panel, and bows upward when over-migrated in the right panel. The slope of the dipping reflector increases with increased velocity, while the flat reflector's slope and vertical position is unaffected.

The creation and application of probabilistic weights to partial images images to create a probabilistic diffraction image is illustrated in Figure 4.4. The front pane of the top left box in Figure 4.4 contains a partial image in velocity generated by stacking the OVC output over slope at 0.5 km, centered directly above the diffractor,

while the side pane shows the slice through this continuation volume for a velocity of 1 km/s, which corresponds to a deterministic image with the true migration velocity.

For this section I calculate the three imaging weights, shown in the bottom row of Figure 4.4. The bottom left box contains W_1, the image semblance calculated according to Equation 4.3. Semblance provides a measure of when slope-gathers are flat, and thus semblance highs can indicate where diffractions have been migrated using the correct velocity. Using this semblance, the expectation velocity, $\bar{v}$, and its variance, σ_v^2, are calculated according to Equations 4.4 and 4.5. A normal distribution is fit to that velocity and variance according to Equation 4.6, generating W_2 which is plotted in the lower middle box.

Notice that $\bar{v}$, the expectation velocity, occurring at the maximum value of W_2, does not track the true velocity, 1 km/s in the shallow or deep portions of the panel. This is because diffraction data does not exist there to be utilized for maximizing semblance. However, $\bar{v} \approx 1$ km/s at 3 s, where the diffraction takes place. The lower right box displays W_3, a weight based on how quickly semblance, or W_1 , changes in space.

The three weights in the lower row of Figure 4.4 posses high values at the time and position of the diffraction at the correct migration velocity but they are also non-zero at other locations where diffraction did not occur. The top middle box of Figure 4.4 shows how multiplying the three weights together further emphasizes the region of the partial image with diffraction data. Multiplying this combined weight by the input $I(t, v, x)$ in the top left box of Figure 4.4 generates the weighted partial

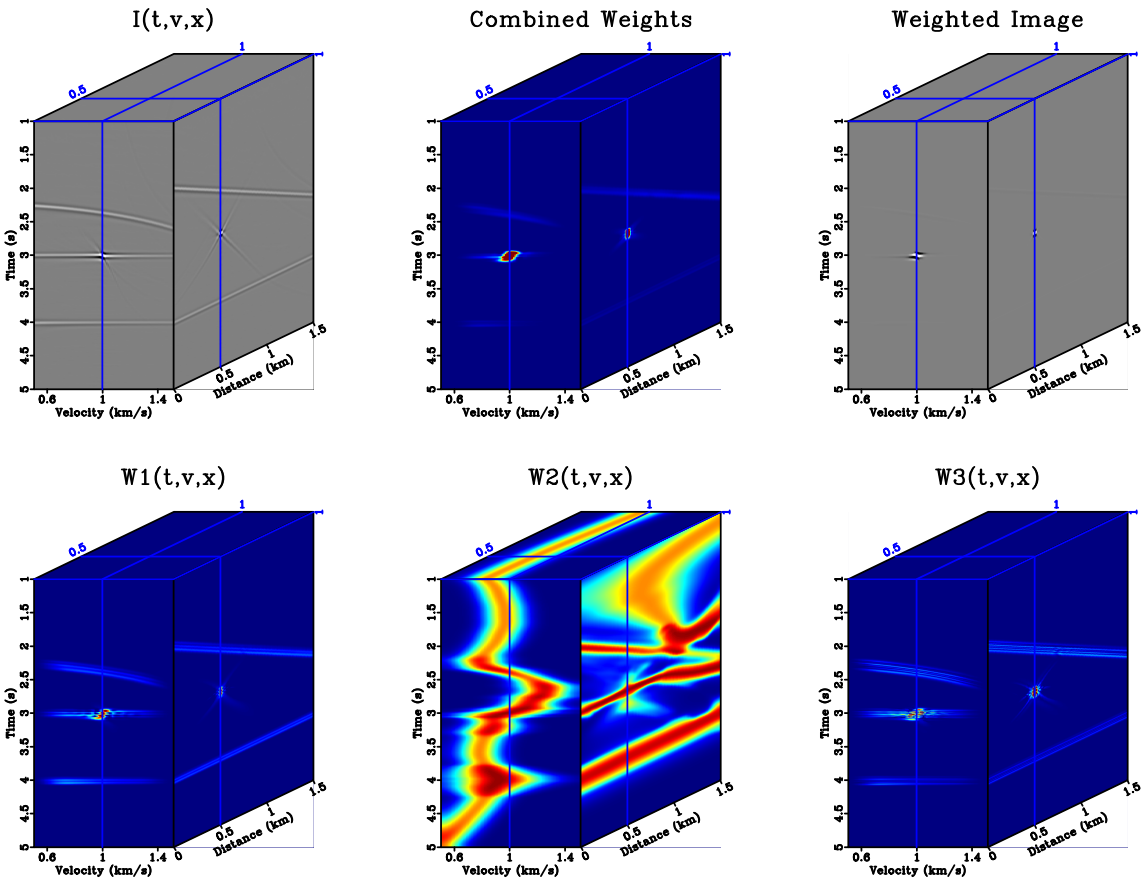


Figure 4.4: Probabilistic migration process illustrating the action of weights on $I(t, v, x)$ for the toy model, with the front panes of box plots centered at $x = 0.5$ km, directly over the diffractor. The top left box contains an output $I(t, v, x)$ volume that results from stacking the output of OVC over slope. Top middle box contains the combined weights that the top left box will be multiplied by. Top right box shows the image on the left multiplied by the combined weights. Bottom left box contains W_1, the image semblance. Bottom middle box contains W_2, a weight normally distributed around expectation velocity using its variance. Bottom right box plots W_3, a weight based on the relative magnitude of semblance's gradient.

image on the top right box of that Figure. Notice how the energy of the diffraction at the correct velocity is emphasized, while other energy present is suppressed.

I generate a suite of images to compare different imaging methods. Figure 4.5(a) contains an "ideal image" generated by convolving the toy model reflectivity with a 10 Hz peak frequency Ricker wavelet. Figure 4.5(b) contains the deterministic image created through migration with the correct velocity. Figure 4.5(c) contains an equal weight comparison image, constructed in the manner of the equal weight images featured in Merzlikin and Fomel (2017). This is the equivalent of stacking the top left box of Figure 4.4 over velocity for each midpoint. Figure 4.5(d) contains our probabilistic diffraction image, generated by stacking the weighted images shown in the top right box of Figure 4.4 over velocity for every midpoint. It is the output of Equation 4.8.

To highlight the wavefield components contributing to each of these images I generate a series of slope gathers centered above the diffractor at 0.5 km. Stacking each of these gathers over slope would create the traces at 0.5 km in their corresponding images. Figure 4.6(a) contains an "ideal gather" generated by warping the ideal image in Figure 4.5(a) to squared time, decomposing it into constituent slope components, and warping the slope decomposed image back to time. The plane wave destruction slope calculated from the ideal image warped to squared time is plotted on this gather with fuchsia x's where reflectors are present. Figure 4.6(b) contains a gather contributing to the deterministic image, Figure 4.5(b). This gather is generated by selecting $I\left(t, x, v_{mig}, p\right)$ for $v_{mig} = 1.0$ km/s. Figure 4.6(c) contains a gather corresponding to the equal weight image, Figure 4.5(c). It is constructed

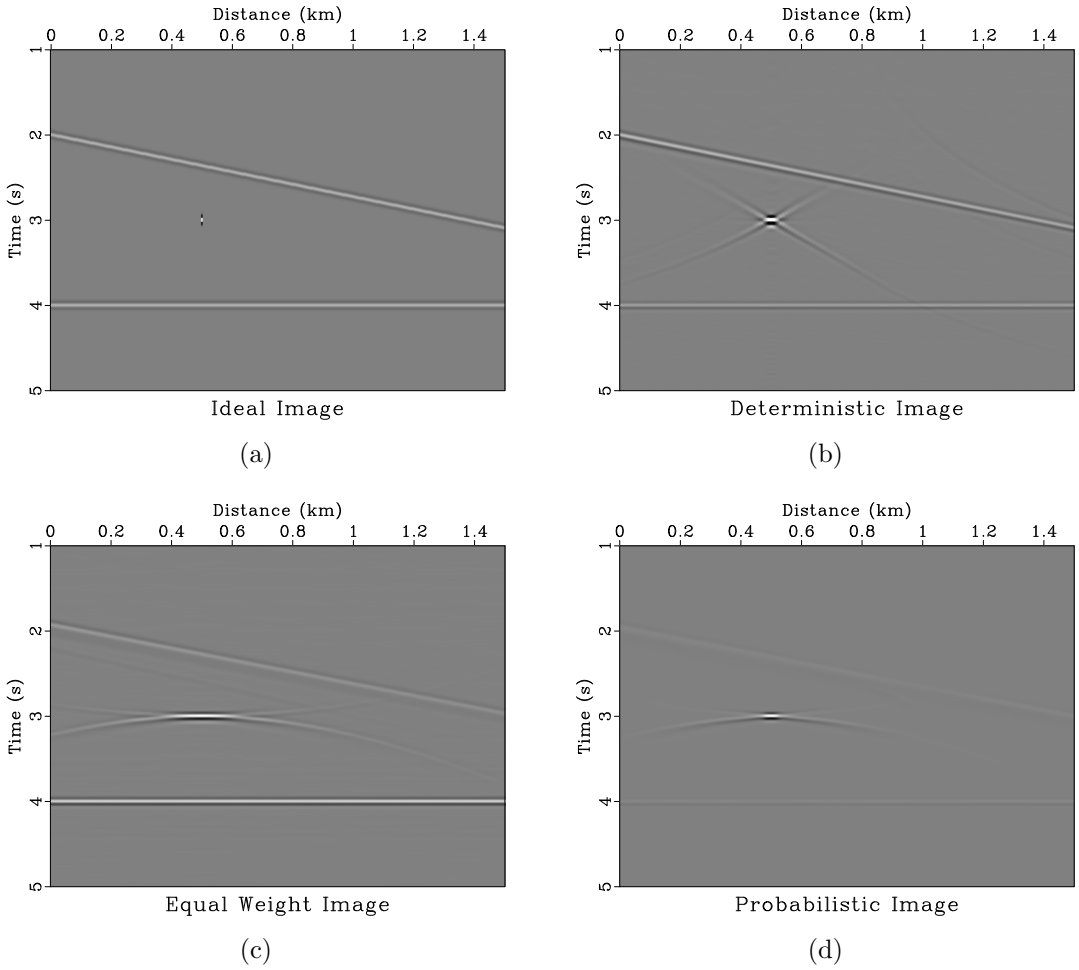


Figure 4.5: Comparison of probabilistic diffraction imaging output with other methods: (a) ideal image corresponding to the toy model reflectivity convolved with a 10 Hz peak frequency Ricker Wavelet; (b) deterministic image created by migrated the data in 4.2 using its migration velocity, 1 km/s; (c) equal weight image generated by stacking $I(t, x, v, p)$ over v and p using equal weights; (d) the probabilistic diffraction image.

by stacking the slope decomposed partial images in velocity,$I\left(t,x,v,p\right)$, over velocity. Figure 4.6(d) contains a gather corresponding to the probabilistic diffraction image, Figure 4.5(d). It is generated by multiplying the combined weights and the slope decomposed partial images in velocity and stacking over velocity.

As one would expect, the ideal image in Figure 4.5(a) features excellently focused and easily discernible diffraction and reflection events. The diffraction in this image is marked by a well defined single point without features radiating away from that point – the point spread function is not present for the diffractor, as the Hessian corresponding to modeling and migration has not been applied to this image. Examining the gather centered at 0.5 km for this image, Figure 4.6(a), the most visible feature is the energetic flat diffraction event extending across all slopes in the gather. One may think of two equivalent reasons for why diffraction energy appears this way in slope gathers. The first is because a point contains information from all slopes – one may see this by taking the fk transform of a dot. The second is that if I think of an idealized diffraction hyperbola extending to infinity, that hyperbola will contain all slopes, from asymptotically vertical up to the right, to flat at the point of the diffractor, to asymptotically vertical downward to the right. Migration with the correct velocity collapses diffraction hyperbolas to a point. Thus a properly migrated idealized diffraction will contain information from all slopes transformed to the same point in space, and such a diffraction in a slope gather immediately above the diffractor would appear as a flat event extending over all slopes. In practice, even in synthetic experiments, seismic data does does not contain such idealized diffraction hyperbolas extending to infinity due to limited spatial geophone or receiver coverage,

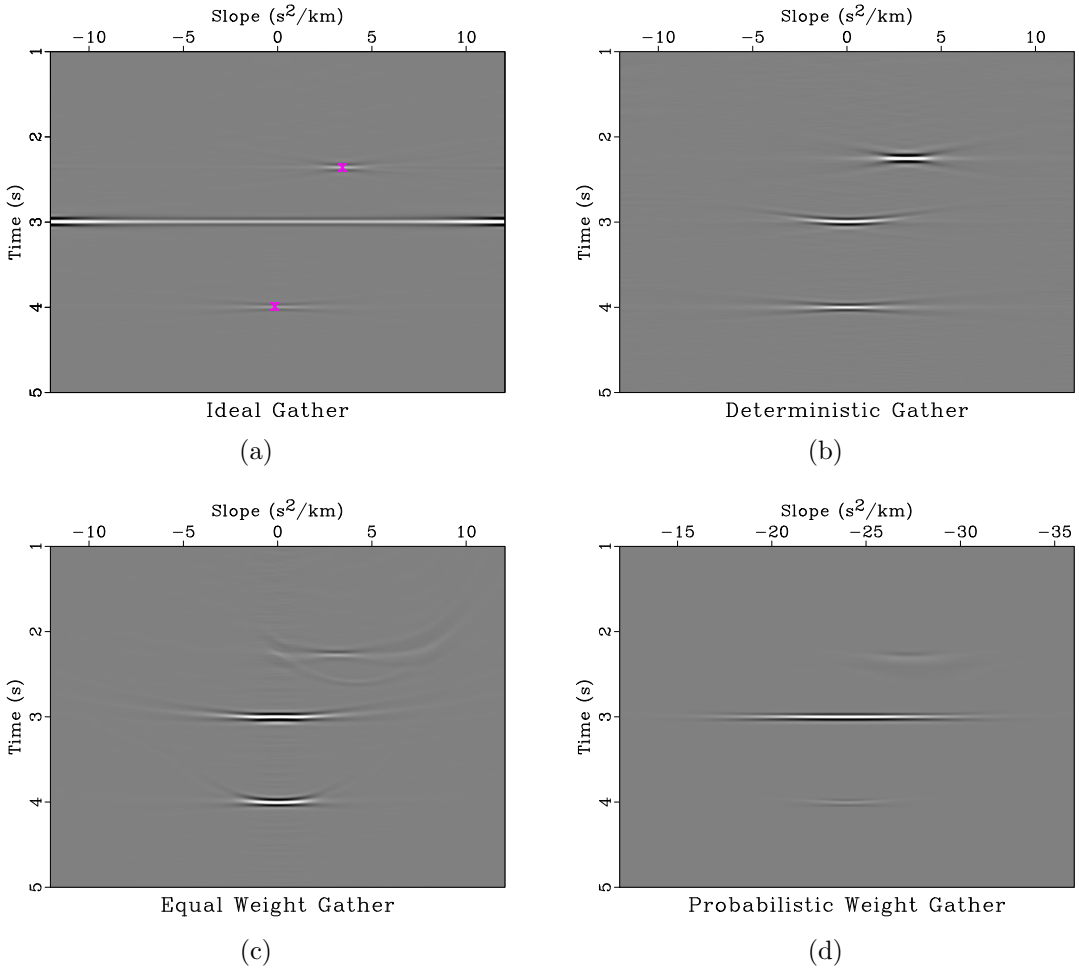


Figure 4.6: Slope gathers centered at 0.5 km, directly above the diffractor, corresponding to the: (a) ideal image, Figure 4.5(a) overlaid by slope denoted using fuchsia x's where reflectors are present; (b) deterministic image, Figure 4.5(b); (c) equal weight image, Figure 4.5(c); (d) the probabilistic diffraction image, Figure 4.5(d).

so only a portion of slopes contribute to the diffraction image. This effect is visible in the point spread function, appearing like a "bow tie". The two reflectors are also visible in this ideal gather, featuring energy confined to a narrower range of slopes centered about the observed slope for each reflector, noted by fuchsia x's. This is intuitive, because planar reflection events contain energy from their dominant slope.

The deterministic image in Figure 4.5(b) appears similar to the ideal image in Figure 4.5(a) with an added "bow tie" point spread function around the diffractor. This result is unsurprising – it assumes complete apriori information about the subsurface velocity field. Differences with the ideal image are caused by the incomplete spatial sampling of the wavefield. Examining the gather corresponding to this image, Figure 4.6(b) shows that as was the case in the ideal gather, three clearly defined events are visible correlating to the two reflections, each centered at their corresponding slope, and the diffraction event. Slope coverage of the diffraction event is more limited in the deterministic gather than in the ideal gather, which is related to the appearance of the point spread function in the stacked image.

In the equal weight image, Figure 4.5(c), the flat reflector is imaged quite well, as all its energy is stationary at the correct time. The dipping reflector appears more smeared because the time its energy achieves the stationary apex in Figure 4.3 changes with velocity, and thus interferes destructively on stacking in this imaging method. The stationary remaining energy corresponds to that of the initial and final velocities, which have no lower or higher velocities respectively to interfere destructively with. This is evident in the gather corresponding to the equal weight image, Figure 4.6(c), where two weak events corresponding to the dipping reflector are visi-

ble. The upper event is related to migration with the initial velocity, and the weaker, lower event is related to migration with the greatest velocity considered. This type of behavior can also be seen in the equal weight diffraction featured in Figure 4.5(c). Rather than having a artifact around the diffraction corresponding to limited spatial coverage, as was the case in Figure 4.5(b), here there is an artifact around the diffraction corresponding to only considering a limited set of migration velocities prior to stacking. Two events, a stronger one bowing downward which is an imprint left by undermigration of diffraction energy with the initial velocity, and a weaker upward bowing event corresponding to diffraction overmigration appear. If I performed this experiment using a dense sampling of velocities spanning all positive numbers, these artifacts would not be present. Diffraction energy is less well spatially resolved than in Figure 4.5(b), appearing more laterally spread out.

The probabilistic diffraction image, Figure 4.5(d) successfully highlights diffraction energy and suppresses energy corresponding to reflectors. A "bow tie" point spread function similar to that appearing in Figure 4.5(b) is not present in this image, although there is a minor artifact appearing like a weaker version of the one surrounding the diffraction in Figure 4.5(c). The diffraction in Figure 4.5(d) has similar lateral resolution to that in Figure 4.5(b) and superior resolution to Figure 4.5(c). Examining the gather corresponding to the probabilistic diffraction image, Figure 4.6(d), shows that although energy corresponding to the two reflection events is present, it is dramatically reduced compared to the deterministic gather in Figure 4.6(b). Diffraction energy in the probabilistic gather spans a greater number of slopes than in the deterministic or equal weight gathers, and appears flat, resembling the diffraction

energy in the "ideal gather" of Figure 4.6(a).

SYNTHETIC EXAMPLE

I illustrate the function of the probabilistic diffraction imaging method and its robustness to noise on a synthetic model consisting of random diffractions in a constant velocity gradient media recreated from Decker et al. (2017a).

Figure 4.7 contains 24 offsets of modeled diffraction data with no added noise. Diffraction data are slope decomposed and OVC is performed over a range of plausible migration velocities. As was the case for the toy model example, I generate partial images and weights as a function of continuation velocity for each midpoint. This process is illustrated for gathers centered below 3.62 km in Figure 4.8. The top left box plot of Figure 4.8 contains partial images in velocity generated by stacking the OVC output, $I(\mathbf{x}, v, p)$ over slope, p. The lower row of Figure 4.8 shows the weights that will be applied to that partial image. The lower left box plot contains W_1, the image semblance calculated according to Equation 4.3. Using this semblance, the expectation velocity, $\bar{v}$, and its variance, σ_v^2, are calculated according to Equations 4.4 and 4.5. A normal distribution is fit to that velocity and variance according to Equation 4.6, generating W_2 as shown in the lower middle box plot. Although $\bar{v}$, the expectation velocity, does not always completely track the true velocity, it is always within one standard deviation. Expectation velocity tends to drift away from the true velocity in the shallow portion of the synthetic, which could be explained by the fact that the shallowest diffractions in the model occur at 0.8 s. Thus, less information is

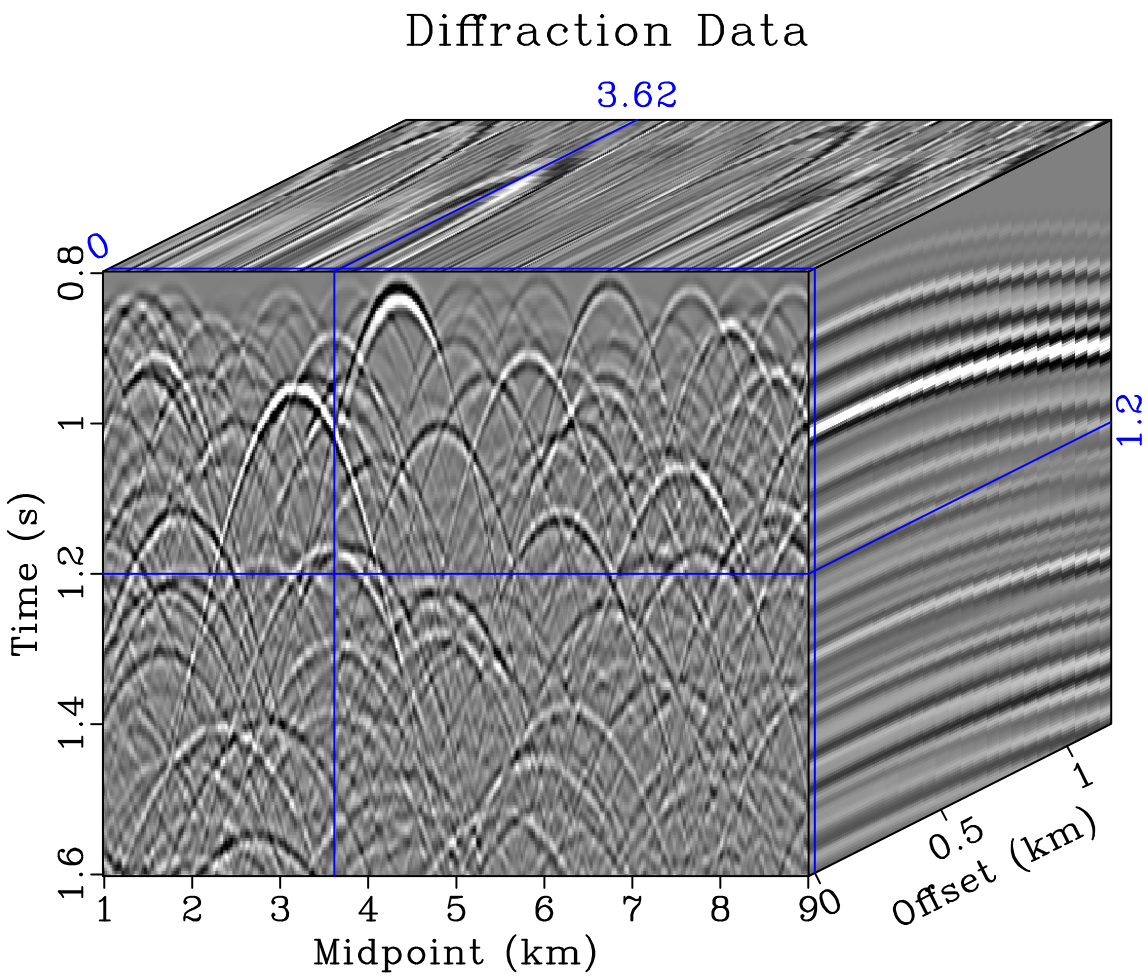


Figure 4.7: Synthetic dataset consisting of diffractions in constant velocity gradient media modeled for 24 offsets with a 50 m increment.

available in the shallow portion to guide the expectation velocity. The right box plot displays W_3, a weight based on how quickly W_1 changes in space.

The three weights are multiplied together to provide the top center box plot of Figure 4.8, the combined weight. Multiplying the partial images in the top left box plot of Figure 4.8 by this combined weight provides the weighted partial images shown on the right box plot.

As was the case with the toy model example used in the methodology section, I generate a suite of images to compare imaging methods on this noiseless synthetic. Figure 4.9(a) shows an ideal image of this synthetic model consisting of the reflectivity model convolved with a 50 Hz peak frequency Ricker wavelet. Figure 4.9(b) contains the deterministic image created using migration with the correct velocity. Figure 4.9(c) contains an equal weight path integral image generated by stacking the partial images in the top left box plot of Figure 4.8 over velocity for each midpoint. Figure 4.9(d) contains our probabilistic diffraction image, generated by stacking the weighted partial images shown in the top right box plot of Figure 4.8 over velocity for every midpoint.

I also create a series of gathers to illustrating the contribution of different wavefield components to the images. Stacking each of these gathers over slope creates the traces at 3.62 km in their corresponding images. Figure 4.10(a) contains an "ideal gather" generated by warping the "ideal image" in Figure 4.9(a) to squared time, decomposing it into constituent slope components, and warping the slope decomposed image back to time. Figure 4.10(b) contains a gather contributing to the deterministic

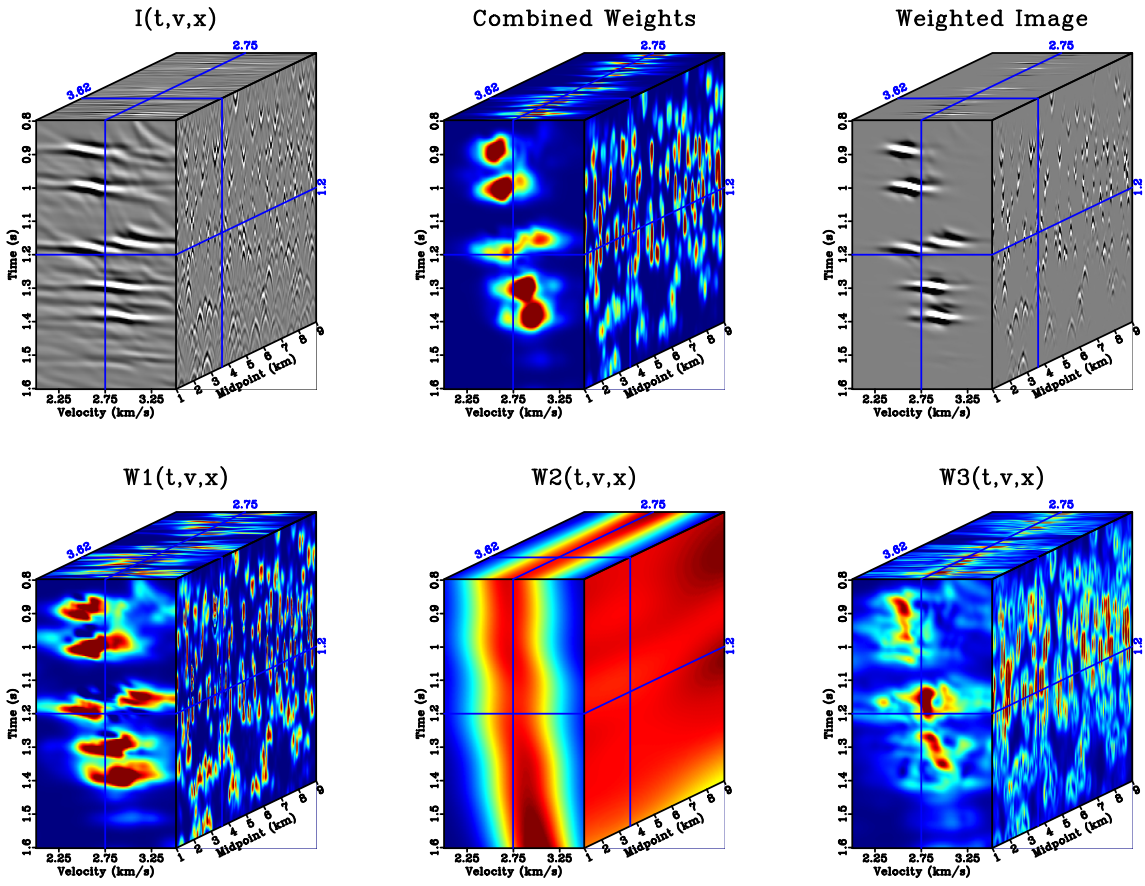


Figure 4.8: Illustration of the probabilistic diffraction imaging process on $I(x, t, v)$ for a gather at centered 3.62 km for the noiseless synthetic. Top left box plot contains partial images output by the OVC process. Top middle box plot contains the combined weights that the left panel will be multiplied by. Top right box plot shows the partial images on the left multiplied by the combined weights. Bottom left box plot shows W_1, the image semblance. Bottom middle box plot contains W_2, a weight normally distributed around expectation velocity, $\bar{v}$, using the expectation velocity's variance, σ_v^2. Bottom right box plot has W_3, a weight based on the relative magnitude of semblance's gradient.

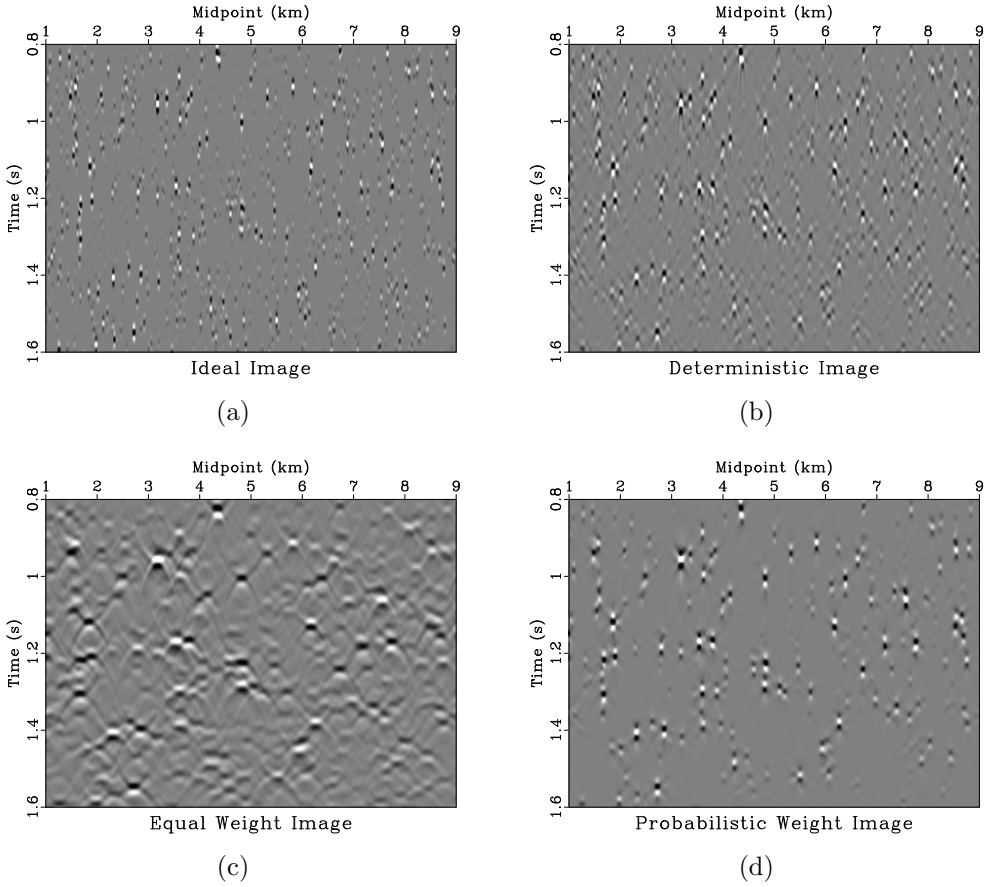


Figure 4.9: Noiseless synthetic model imaging comparison: (a) ideal image consisting of synthetic model reflectivity convolved with a 50 Hz peak frequency Ricker wavelet; (b) deterministic image created by migrating the noiseless synthetic data using its true velocity; (c) equal weight path integral image generated by stacking $I(t, x, v, p)$ over v and p using equal weights; (d) probabilistic diffraction image generated by stacking $I(t, x, v, p)$ over v and p using probability weights.

image, Figure 4.9(b). This gather is generated by selecting $I(t, x, v_{mig}, p)$ for the true migration velocity. Figure 4.10(c) contains a gather corresponding to the equal weight image, Figure 4.9(c). It is constructed by stacking the slope decomposed partial images in velocity, $I(t, x, v, p)$ over velocity. Figure 4.10(d) contains a gather corresponding to the probabilistic diffraction image, Figure 4.9(d). It is generated by multiplying the combined weights and the slope decomposed partial images in velocity and stacking over velocity.

In this noiseless example, the deterministic image in Figure 4.9(b) is unsurprisingly able to image most of the diffractions in the ideal image, Figure 4.9(a). This is because in this example, the deterministic imaging process receives perfect apriori knowledge of the subsurface velocity field, so it would be expected to be extremely effective at imaging diffractions. Notice that in the process of modelling and migration, the dynamic range of diffractions appears to be amplified in the deterministic image relative to the ideal image, which consists of the reflectivity model convolved with a wavelet. Diffractions which are weak in the ideal image appear weaker in the deterministic image, and strong diffractions appear stronger. An example of this behavior is in the cluster of weak diffractions around 1.55 s between 3 and 4 km, which are relatively more energetic and better resolved in Figure 4.9(a) than Figure 4.9(b). The deterministic image also has slightly poorer lateral resolution of diffractions than the ideal image. The deterministic gather, Figure 4.10(b) appears similar to the ideal gather in Figure 4.10(a)k although energy is primarily confined to slopes between -1 and 1 s^2/km. Energy in these gathers that is flat corresponds to diffractions immediately below the gather. Energy that is sloping correlates to diffractions away from

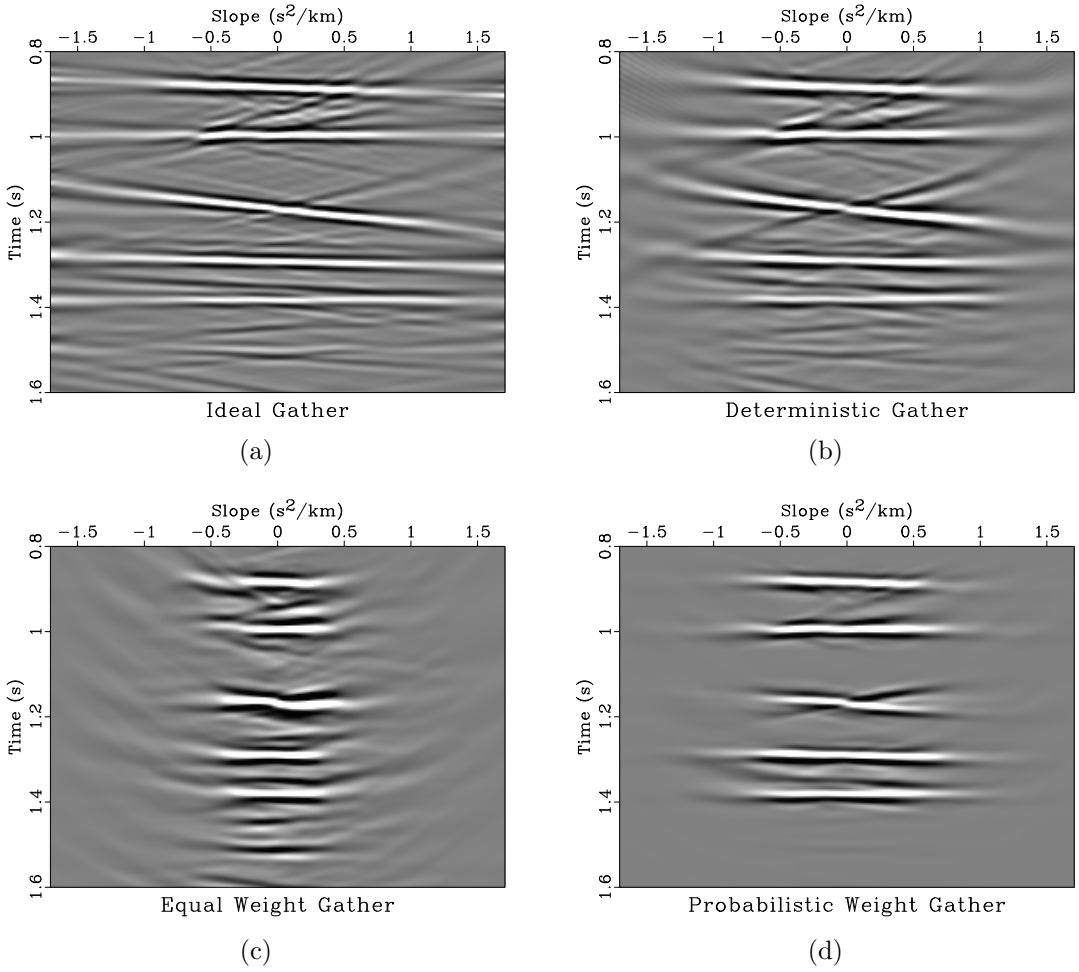


Figure 4.10: Slope gathers centered at 3.62 km for the noiseless synthetic corresponding to the: (a) ideal image, Figure 4.9(a); (b) deterministic image, Figure 4.9(b); (c) equal weight image, Figure 4.9(c); (d) the probabilistic diffraction image, Figure 4.9(d).

the gather's midpoint. The greater the slope of energy in the gather, the further away the event is. For example, the sloping event centered around 1.15 s in the ideal gather, Figure 4.9(a), corresponds to a strong diffraction event occurring around 1.15 s below 3.5 km, slightly to the left of this gather which is centered at 3.62 km. Only diffraction events events that are nearly flat in these gathers will be stationary upon stacking. The sloping events will interfere destructively with themselves, as is the case with the sloping event centered around 1.15 s, which does not appear in the ideal image at 3.62 km, although some energy is present there in the deterministic image, Figure 4.9(b) because diffractions in that image are less spatially focused and thus wider.

The equal weight image, Figure 4.9(c), resolves many of the diffractions visible in the ideal image, Figure 4.9(a) but with significantly reduced spatial resolution. The increase in dynamic range visible in the deterministic image, Figure 4.9(b) is also visible here, and often diffractions which are separate in the ideal image appear merged in the equal weight image. It is often difficult to discern individual weak diffraction events in this image. Examining the equal weight gather, Figure 4.10(c), notice that the range of slopes where energy is present is more limited, typically ranging between -0.5 and 0.5 s^2/km. Much of the sloping energy associated with diffractions centered beneath different midpoints which is visible in Figures 4.10(a) and 4.10(b) is absent here. Because the equal weight image has poorer lateral resolution than the ideal or deterministic images, several diffraction events which are centered nearby appear as flat diffraction energy in the equal weight gather, but appear as sloping events in the deterministic or equal weight gathers. Two examples of this occur at approximately

1.45 and 1.55 s. Note that although the equal weight image does not resolve diffraction events as well as the deterministic image, it was not the recipient of apriori knowledge of the subsurface velocity field during the imaging process.

The probabilistic diffraction image, Figure 4.9(d) effectively focuses the energy of moderate to strong diffractions. Diffractions seen in the probabilistic image have better spatial resolution than those in the equal weight image. Although some weak diffractions that are present in the ideal image of Figure 4.9(a) and the deterministic image of Figure 4.9(b), recall that unlike the deterministic image, the probabilistic image did not receive apriori knowledge of the subsurface velocity field to use during imaging. Instead, such a field was output during the creation of probability weights. Examining the gather corresponding to the probabilistic image, Figure 4.10(d), notice that the gather has a cleaner appearance. Energy present in the probabilistic gather spans a similar range of slopes to that of the deterministic gather, roughly -1 to 1 s^2/km. Events visible in the probabilistic weight gather tend to be the flat ones indicative of a diffraction at that location. Sloping events corresponding to diffractions at other midpoints tend to be suppressed. This is because the weights are primarily based on semblance, a measure of gather coherence. Thus, the method amplifies the nearly flat coherent events possessing high semblance values. The use of weights tied to coherence explains why some weak diffractions can be suppressed in the imaging process. This synthetic model possesses dense field of diffractions, as is visible in the diffraction data of Figure 4.7. The moveout curves associated with these diffractions frequently intersect, which leads to the phenomenon of sloping energy corresponding to diffractions centered at a different midpoint intersecting the flat energy in a gather

corresponding to a diffraction located at that midpoint. The resulting superposition of sloping and flat energy is less laterally coherent, and thus will have a lower semblance value than if the sloping energy was not present. The effect becomes stronger when the coherent flat event is similarly or less energetic to intersecting events. This behavior can be seen in gently sloping events below 1.4 s in the four gathers. Although these events slope more gently than the event at 1.15 s, their superposition is laterally incoherent. Because the probabilistic imaging process amplifies events that are laterally coherent in these gathers, which is a typical feature of diffractions, they are dramatically suppressed.

I repeat the synthetic experiment adding 10-60 Hz band passed Gaussian noise to the diffraction data following modeling to show the robustness of the probabilistic imaging method to noise. The added noise is strong, it has a RMS value approximately 16.5 times greater than that of the noiseless diffraction data from the initial part of this experiment. Noisy zero offset data are shown in Figure 4.11. I perform OVC and again illustrate the probabilistic imaging process in Figure 4.12. The top left box plot of Figure 4.12 contains partial images output by OVC after stacking over slope. The slope-decomposed images output by OVC are also used to calculate the weight functions shown in the bottom row of Figure 4.12. In order to suppress unreasonable values, tapered muting is applied to W_1 before calculating the other weights. The three weights are multiplied together to create the combined weight shown in the top middle box plot of Figure 4.12. Multiplying the combined weight by the partial images in the left box plot creates the weighted partial images in the top right box plot of Figure 4.12.

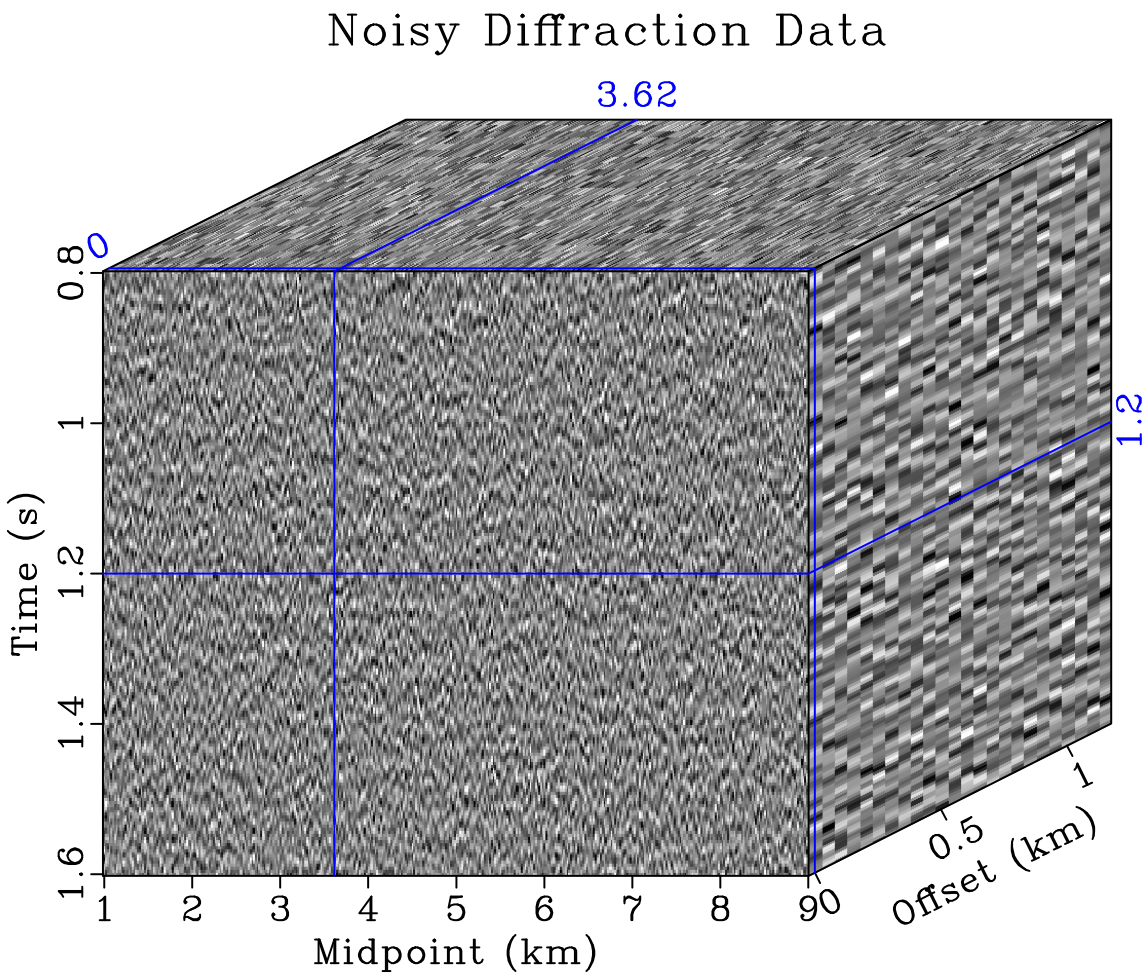


Figure 4.11: Synthetic dataset consisting of diffractions in constant velocity gradient media modeled for 24 offsets with a 50 m increment from Figure 4.7 with band passed Gaussian noise added.

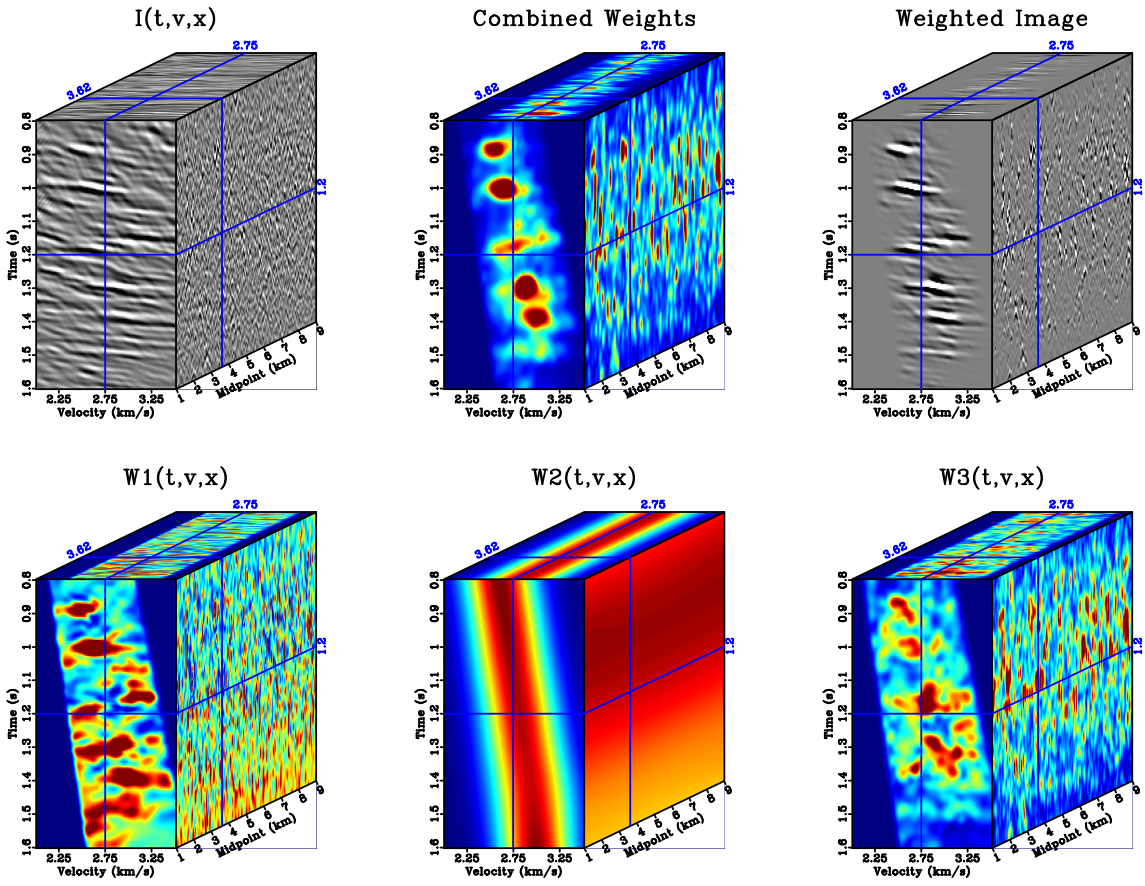


Figure 4.12: Illustration of the probabilistic diffraction imaging process on $I(x, t, v)$ for a gather at centered 3.62 km for the noisy synthetic. Left box plot contains partial images output by the OVC process. Central box plot contains the combined weights that the left panel will be multiplied by. Right box plot shows the partial images on the left multiplied by the combined weights. Bottom left box plot shows W_1, the image semblance. Bottom middle box plot contains W_2, a weight normally distributed around expectation velocity, $\bar{v}$, using the expectation velocity's variance, σ_v^2. Bottom right box plot has W_3, a weight based on the relative magnitude of semblance's gradient.

As with the noiseless example, I generate a suite of images to compare imaging methods. Figure 4.13(a) shows an ideal image of this synthetic model consisting of the reflectivity model convolved with a 50 Hz peak frequency Ricker wavelet. This image is the same as that in Figure 4.9(a) as I used the same model and wavelet. Figure 4.13(b) contains the deterministic image created using migration with the correct velocity. Figure 4.13(c) contains an equal weight path integral image generated by stacking the partial images in the top left box plot of Figure 4.12 over velocity for each midpoint. Figure 4.13(d) contains our probabilistic diffraction image, generated by stacking the weighted partial images shown in the top right box plot of Figure 4.12 over velocity for every midpoint.

I similarly create a series of gathers to illustrating the contribution of different wavefield components to the noisy images at the same midpoint location as the gathers in the noiseless example, 3.62 km. Figure 4.14(a) contains an ideal gather generated by warping the ideal image in Figure 4.13(a) to squared time, decomposing it into constituent slope components, and warping the slope decomposed image back to time. As was the case with the ideal image, this ideal gather is the same as that in Figure 4.9(a). Figure 4.14(b) contains a gather contributing to the deterministic image, Figure 4.13(b). This gather is generated by selecting $I\left(t, x, v_{mig}, p\right)$ for the true migration velocity. Figure 4.14(c) contains a gather corresponding to the equal weight image, Figure 4.13(c). It is constructed by stacking the slope decomposed partial images in velocity, $I\left(t, x, v, p\right)$ over velocity. Figure 4.14(d) contains a gather corresponding to the probabilistic diffraction image, Figure 4.13(d). It is generated by multiplying the combined weights and the slope decomposed partial images in

velocity and stacking over velocity.

The noisy deterministic image, Figure 4.13(b), contains significantly more noise than its noiseless counterpart, Figure 4.9(b). Although some of the strongest diffractions seen in the ideal image, Figure 4.13(a) can be seen in the deterministic image, it can be difficult to tell which objects are diffractions and which are noise without the ideal image as reference. The corresponding deterministic gather, Figure 4.14(b) is also noisy when compared to the ideal gather, Figure 4.14(a). As was the case with the deterministic image, although most of the planar diffraction events from the ideal gather can be seen in the deterministic gather and identified because of their lateral coherence, their amplitudes are typically not much stronger than those of background noise.

The noisy equal weight image, Figure 4.13(c), is corrupted by noise. Few diffractions can be discerned in the image, and it is difficult to say with certainty what features are diffractions and what are simply high amplitude noise. This behavior is reflected in the corresponding equal weight gather, Figure 4.14(c), which contains numerous high amplitude events not present in the noiseless equal weight gather, Figure 4.10(c).

The probabilistic diffraction image shown in Figure 4.13(d) is resilient to noise, featuring most of the diffractions visible in Figure 4.9(d). Background noise, though present, is not as powerful as the strong diffractions, and is significantly suppressed relative to the noisy deterministic and equal weight images. Features in the probabilistic weight gather, Figure 4.14(d) appear more laterally coherent and less noisy

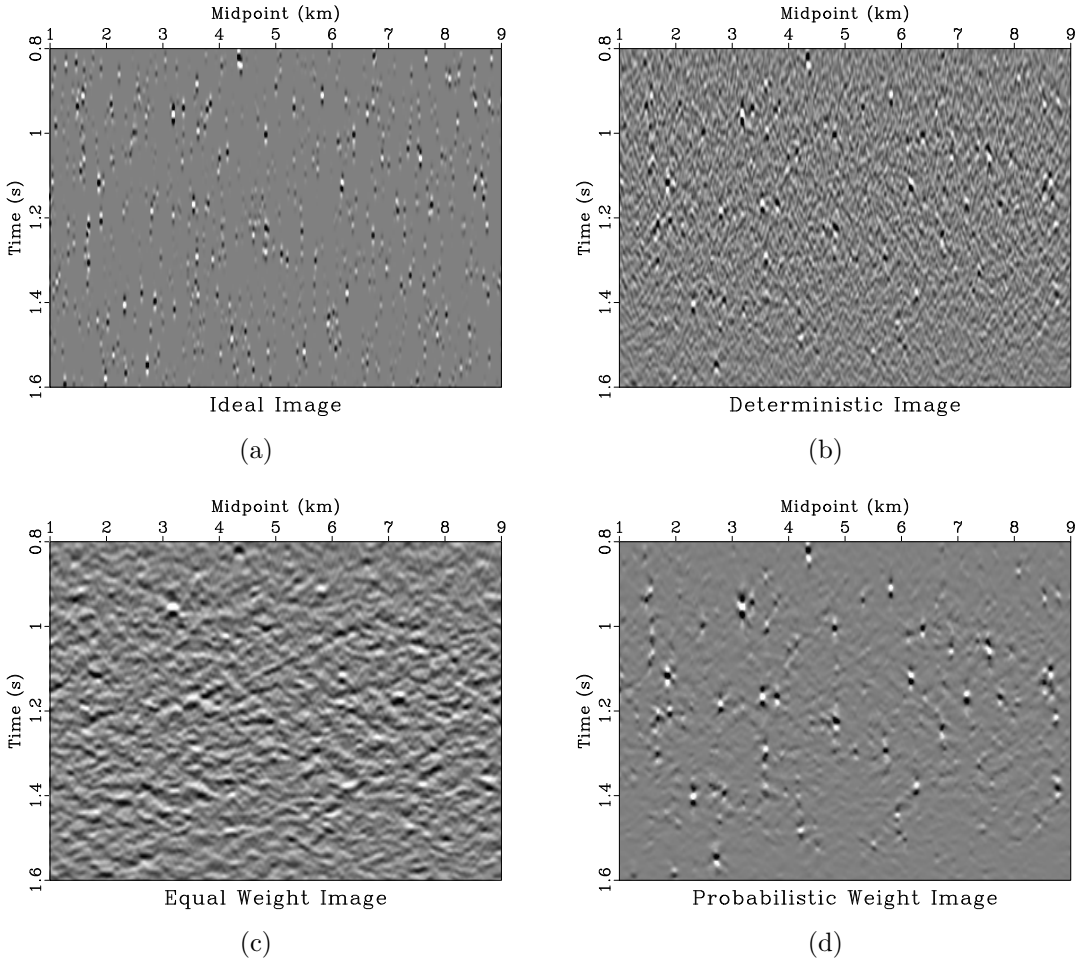


Figure 4.13: Noisy synthetic model imaging comparison: (a) ideal image consisting of synthetic model reflectivity convolved with a 50 Hz peak frequency Ricker wavelet; (b) deterministic image created by migrating the noiseless synthetic data using its true velocity; (c) equal weight path integral image generated by stacking $I(t, x, v, p)$ over v and p using equal weights; (d) probabilistic diffraction image generated by stacking $I(t, x, v, p)$ over v and p using probability weights.

than those in the deterministic or equal weight gathers, signifying that the energy present is far more likely to be caused by diffraction. Indeed, comparing this noisy probabilistic weight gather to the noiseless one in Figure 4.10(d) shows that most events appearing in the noisy gather have corresponding events in the noiseless one. Also note that the noisy deterministic image received perfect apriori knowledge of underlying velocity to use in the migration process, while this probabilistic image did not.

FIELD DATA EXAMPLE

I apply the method to a 2D deepwater field dataset acquired to image the Nankai Trough subduction zone in Japan off the shore of Honshū in the Philipine Sea, where the Philipine Plate subducts below the Eurasian Plate. Moore et al. (1990) contains relevant data acquisition parameters and processing results associated with this dataset, which is referred to as NT62-8. Moore and Shipley (1993) performed structural interpretation on the line. Additional regional context may be found in Moore et al. (2007) and Bangs et al. (2009). This experiment examines CMPS 900-1301 from that line, previously used by Forel et al. (2005) and Decker et al. (2017a). This portion of the line highlights the transition from trench, stretching to to the south, or off to the left of our study area with lower CMP numbers, to a highly deformed accretionary wedge of sedimentary rocks featuring numerous thrust faults that overlays the subducting oceanic crust within our study area. This dataset features numerous diffractions, making it well suited for application to diffraction imaging.

Data are pre-processed by correcting traces to be zero-mean, using a 10-125

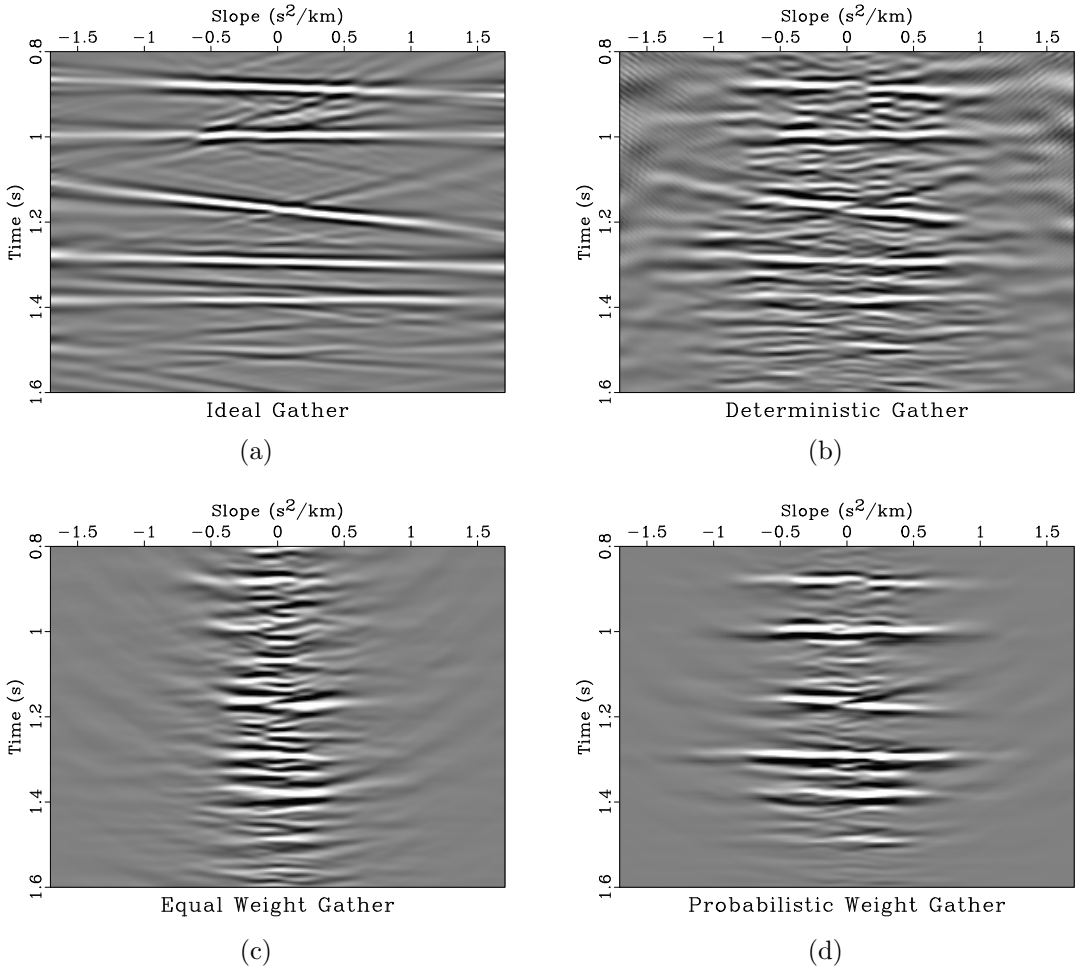
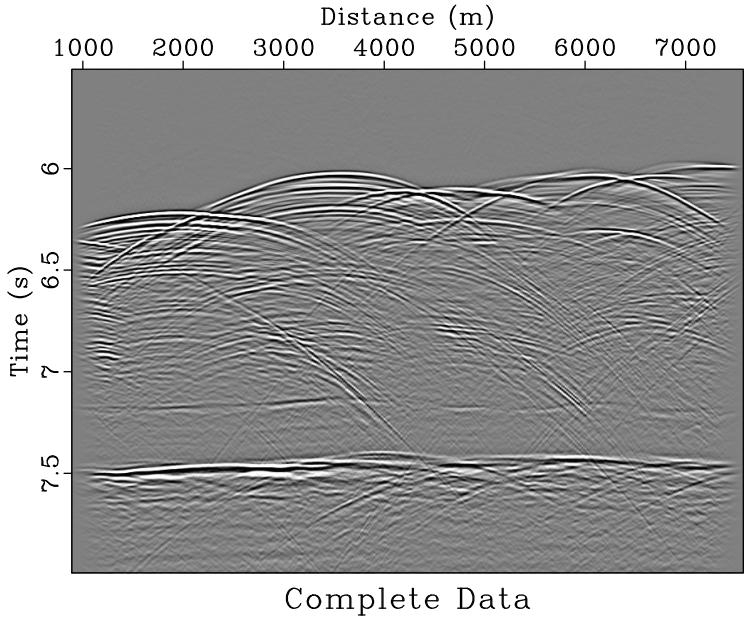
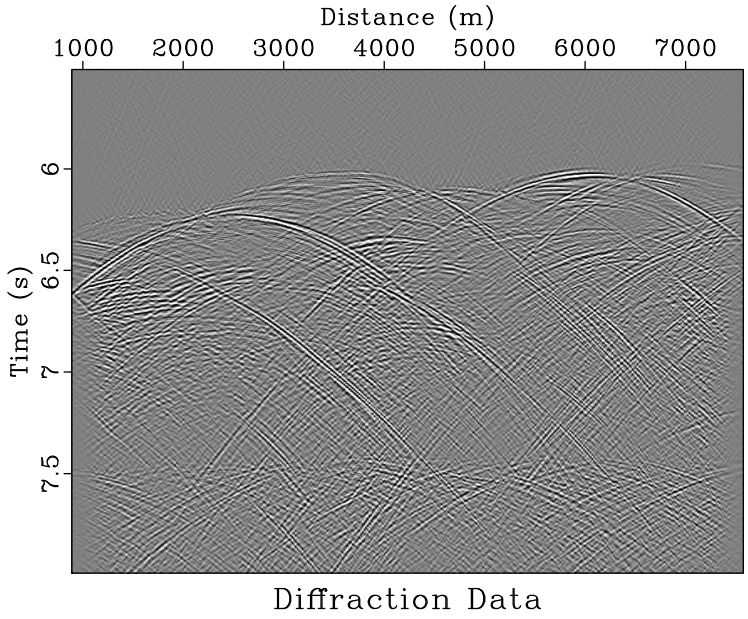


Figure 4.14: Slope gathers centered at 3.62 km, directly above the diffractior corresponding to: (a) ideal image, Figure 4.13(a); (b) deterministic image, Figure 4.13(b); (c) equal weight image, Figure 4.13(c); (d) the probabilistic diffraction image, Figure 4.13(d).



(a)



(b)

Figure 4.15: Field data from Nankai Trough: (a) complete data; (b) plane-wave destruction diffraction data.

Hz bandpass filter, applying surface consistent amplitude correction, and resampling to 4ms. Data are then DMO stacked, shown in Figure 4.15(a), and diffractions are extracted through plane-wave destruction (Fomel et al., 2007). Diffraction data are displayed in Figure 4.15(b). OVC is performed on the diffraction data from Figure 4.15(b), outputting a suite of slope-decomposed diffraction images for 60 different migration velocities, beginning at 1.4 km/s with a 20 m/s increment.

Slope-decomposed images are used to perform the probabilistic diffraction imaging process, illustrated in Figure 4.16. Stacking the slope-decomposed partial images, $I(t, v, x, p)$ over slope p provides the partial images in the top left box plot of Figure 4.16. Semblance is calculated from slope-decomposed partial images according to Equation 4.2, and is used to generate the imaging weights shown in the bottom row of Figure 4.16 as well as the expectation velocity and its variance, plotted in Figures 4.17(a) and 4.17(b) respectively.

Multiplying the weights together provides the combined weight in the top middle box plot of Figure 4.16. Multiplying those weights by the partial images in velocity in the top left box plot generates the weighted partial images in the top right box plot. I use the expectation velocity in Figure 4.17(a) to generate a deterministic complete image, shown in Figure 4.18(a). I generate a deterministic diffraction image plotted in Figure 4.18(b) by migrating the diffraction data in Figure 4.15(b) using the expectation velocity in Figure 4.17(a), and a equal weight diffraction image, Figure 4.18(c) by stacking the partial images in the upper left box plot of Figure 4.16 over velocity with equal weights. Stacking the weighted partial images in the upper right panel of Figure 4.16 over velocity provides the probabilistic diffraction image in

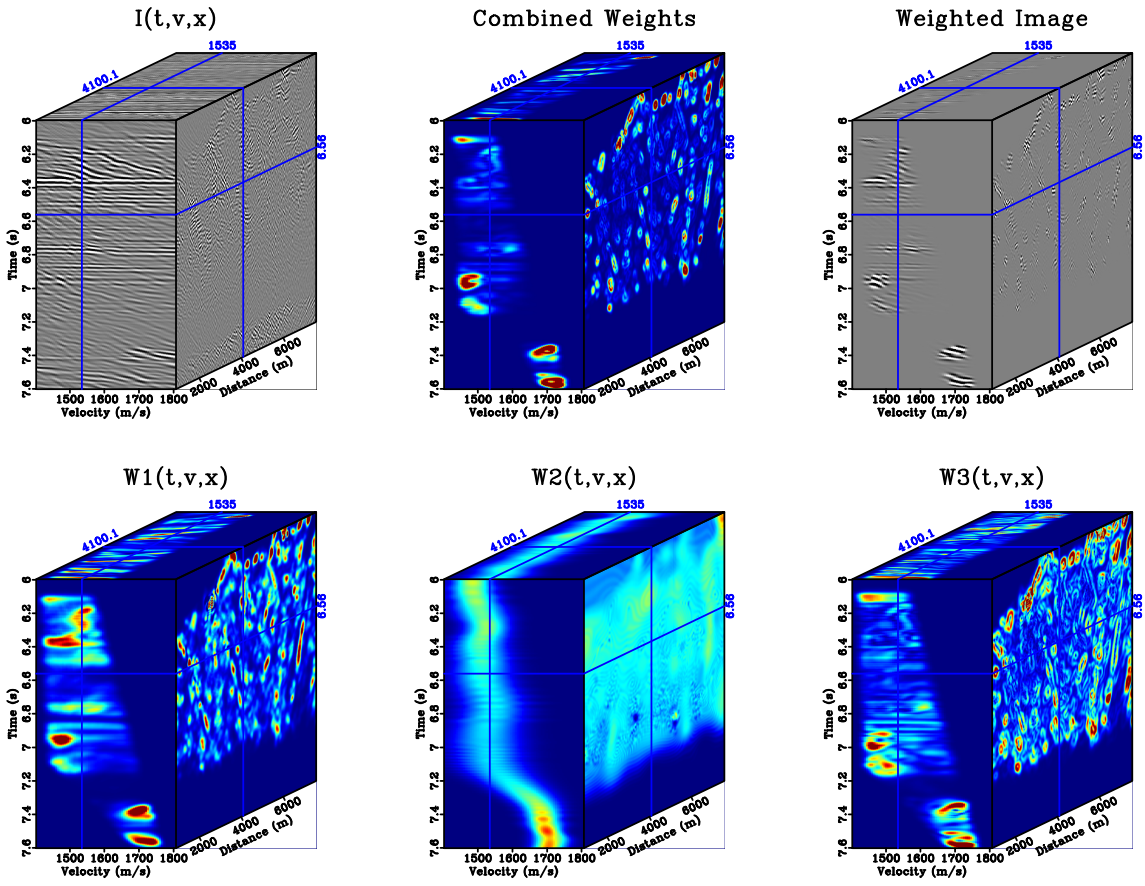
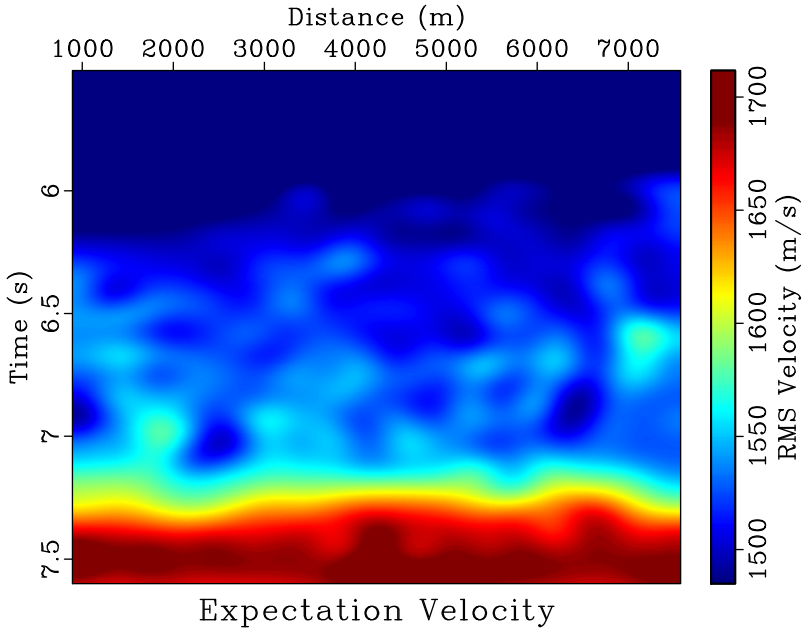
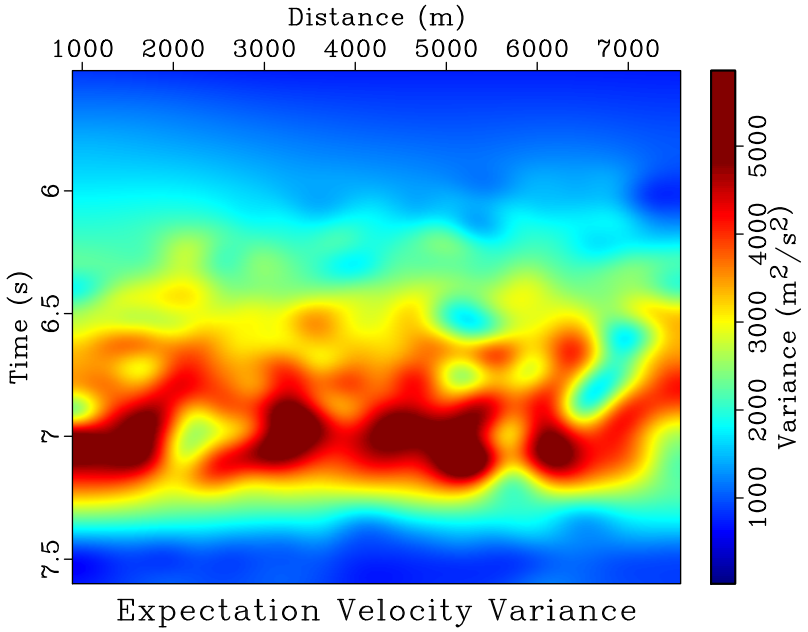


Figure 4.16: Illustration of the probabilistic diffraction imaging process on the Nankai field dataset on a gather centered at 4100 m: Top left box plot contains partial images output by the OVC process after stacking over slope. Top middle box plot contains the combined weights that the left panel will be multiplied by. Top right box plot shows the partial images on the left multiplied by the combined weights. Bottom left box plot shows W_1, the image semblance. Bottom middle box plot contains W_2, a weight normally distributed around expectation velocity, $\bar{v}$, using the expectation velocity's variance, σ_v^2. Bottom right box plot has W_3, a weight based on the magnitude of semblance's gradient.



(a)



(b)

Figure 4.17: Nankai Trough velocity attributes calculated from semblance: (a) expectation velocity; (b) velocity variance.

Figure 4.18(d).

I generate a suite of slope gathers for midpoints at 4100 m to illustrate how different wavefield components contribute to the four images. Figure 4.19(a) contains a slope gather corresponding to the Nankai Trough complete image, Figure 4.18(a). That gather is overlaid by a fuchsia plot of image slope at that midpoint computed by warping the complete image to squared time, determining slope, and then warping that slope back to non-squared time following the same process used in the toy model example to find the slope plotted in Figure 4.18(a). Figure 4.19(b) displays the slope gather corresponding to the deterministic diffraction image, Figure 4.18(b), Figure 4.19(c) features the gather corresponding to the equal weight image, Figure 4.18(c), and Figure 4.19(d) has the gather corresponding to the probabilistic diffraction image, Figure 4.18(d).

The probabilistic weight diffraction image better highlights diffractive features and suppresses remnant reflection signal than the deterministic or equal weight diffraction images. Diffractions delineating the seafloor between about 5.9 and 6.3 s, and likely tied to slumps, are well resolved in the probabilistic diffraction image, but less defined in the deterministic and equal weight images where they are either less focused or overwhelmed by their point spread functions and noise, making it more difficult to discern the location of the seafloor.

Diffractions highlighting thrust faults, particularly two that intersect the seafloor at 1000 km and 2300 km and extend downward to the right to a relatively flat décollement extending laterally at about 7.2 s and defining the plate boundary, are

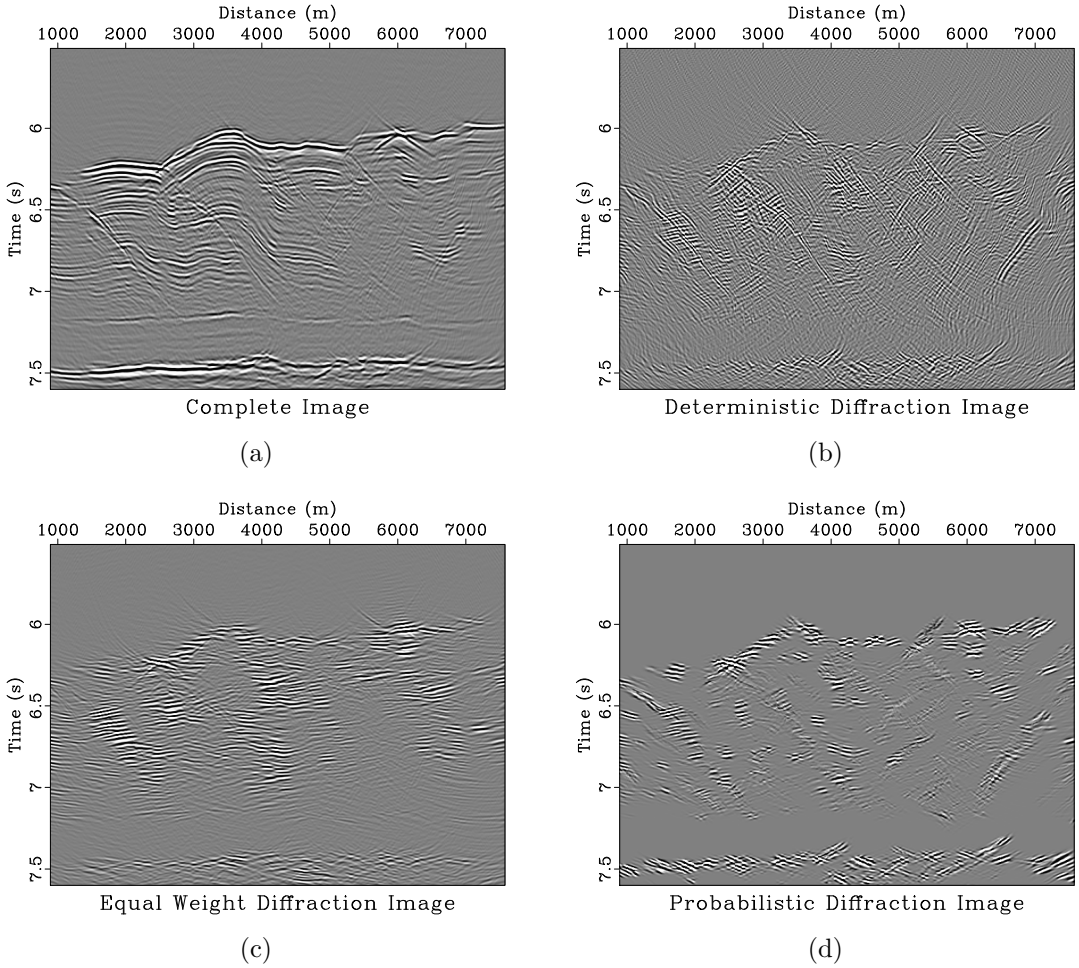


Figure 4.18: Images of the Nankai Trough: (a) deterministic complete image generated by migrating the complete data in Figure 4.15(a) using the expectation velocity, Figure 4.17(a); (b) deterministic diffraction image generated by migrating the diffraction data, Figure 4.15(b), using the expectation velocity, Figure 4.17(a); (c) diffraction image generated through the equal weight stack over velocity of the the partial images in the left box plot of Figure 4.16; (d) probabilistic weight diffraction image created by stacking the weighted partial images in the top right box plot of Figure 4.16 over velocity.

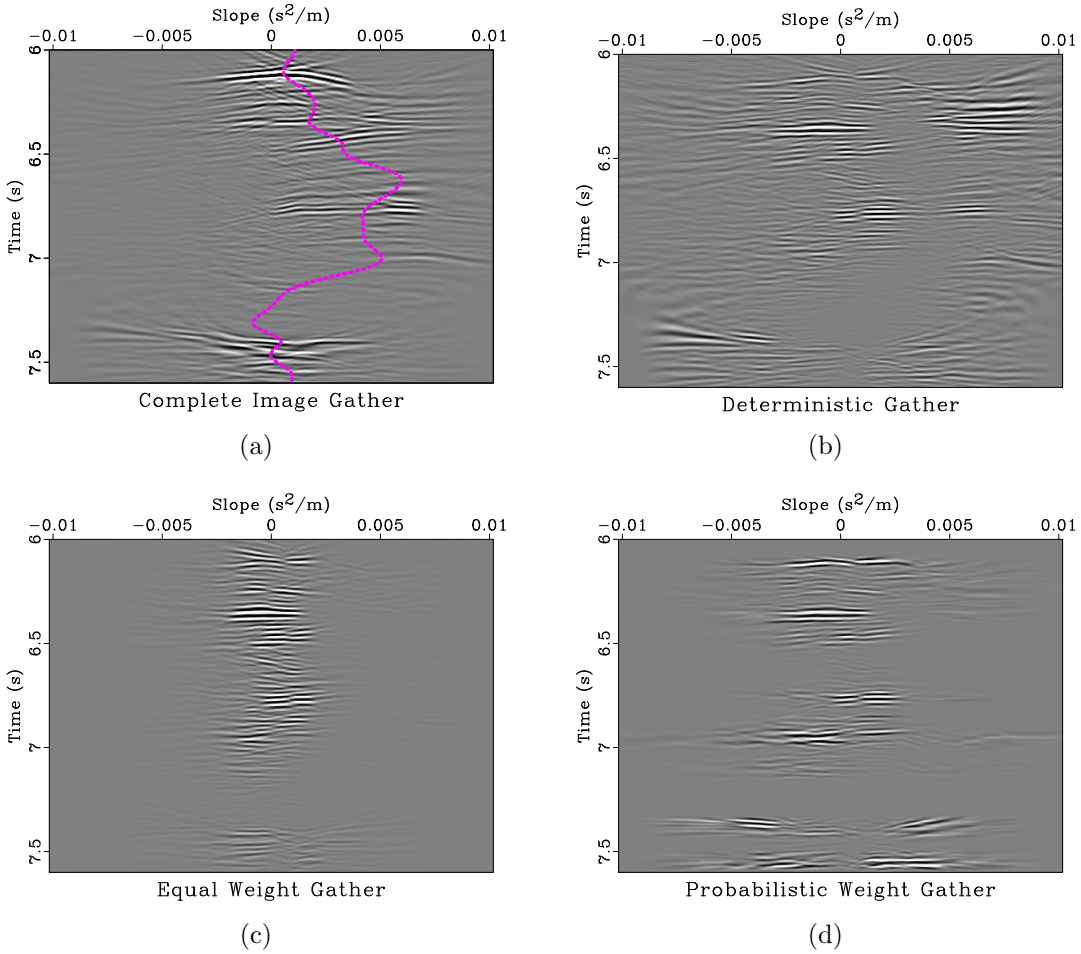


Figure 4.19: Slope gathers centered at 4100 m corresponding to (a) the complete image, Figure 4.18(a); (b) the deterministic diffraction image, Figure 4.18(b); (c) the equal weight image, Figure 4.18(c); (d) the probabilistic diffraction image, Figure 4.18(d).

clearly visible in the probabilistic diffraction image. These faults are more difficult do distinguish in the deterministic diffraction image due to remnant reflections and in the equal weight diffraction image due to background noise.

A highly diffractive layer extending laterally at about 7.5 s marking the transition from sedimentary to crystalline rock is more localized in the probabilistic diffraction image than the deterministic or path integral diffraction images. This boundary marks the deepest diffractions observed in this data, so velocity data below this boundary corresponding to the crystalline rock is unavailable, as diffraction energy is not observed traveling through it.

In general the deterministic and equal weight diffraction images contain more background noise than the probabilistic image, although all three methods are able to isolate weak diffractions in the décollement at about 7.2 s and identify the lowest sedimentary region between the décollement and crystalline rock transition at 7.5 s as relatively free of diffraction energy.

Examining the gathers centered at 4100 m, notice that much of the energy present in the complete image slope gather, Figure 4.19(a), surrounds the dominant slope denoted by the dashed fuchsia line. This is expected, as reflection signal is often the most powerful feature of a seismic image. The stationary reflection energy surrounding that line is not present in the other gathers, as it has been mostly removed by plane wave destruction – notice in the deterministic gather, Figure 4.19(b), that energy around the dominant slope is significantly suppressed. Although not a central argument in this paper, this is interesting because it illustrates how the deterministic

diffraction imaging workflow employed in the creation of this image, which was based on reflection removal by removing energy possessing the dominant slope in the data domain as identified by plane-wave destruction filters following Fomel (2002), has a similarity to the diffraction imaging method of Moser and Howard (2008), which involves directly applying a mute around the dominant slope in gathers similar to these to surpress stationary reflection energy. In effect, the plane-wave destruction process has generated similarly muted data when viewed in these gathers. A key difference, is that plane-wave destruction is able to at least partially differentiate between diffraction energy underlying reflection energy rather than completely masking it. This feature is particularly apparent in the two gathers when dominant slopes are near zero, as can be seen near 6.1 s and 7.4 s where energy remains near the dominant slope value in the deterministic gather.

The three diffraction image gathers, the deterministic gather of Figure 4.19(b), the equal weight gather of Figure 4.19(c), and the probabilistic gather of Figure 4.19(d), have significantly different appearance. Although energy near the complete image gather's dominant slope tends to be suppressed in all three gathers, as one would expect from diffraction images, the ranges over which energy is present differs significantly. Most energy in the equal weight gather is confined within ± 0.003 s^2/m, while that of the other two gathers has a larger range, with at least some energy extending the width of each gather. The probabilistic weight gather has a more sparse, coherent, and clean appearance when compared to the other two, while the deterministic gather features more chaotic, less coherent features. The deterministic gather also features some interesting coherent energy isolated to slope values with magnitudes

greater than 0.005 s^2/m which is likely tied to out of plane diffractions, reflection energy that does not have the dominant slope at a location, or migration artifacts. An example of such an feature occurs in the deterministic gather between 6.2 and 6.4 s, which corresponds to events in the deterministic diffraction image, Figure 4.18(b), sloping steeply downward to the right at 4100 m between 6.2 and 6.4 s. In this case they appear to be related to reflections off of a dipping fault interface cutting through the strata which define dominant slope. Although these features are interesting and can be useful for identifying faults, they are not diffractions. Also notice that within the three diffraction gathers, coherent energy is not present between 6.5 and 6.7 s. Indeed, in the probabilistic gather very little energy at all is present in that interval, and thus the probabilistic diffraction image, Figure 4.18(d) is mostly blank in that interval at 4100 m. However, in both the deterministic diffraction image, Figure 4.18(b), and the equal weight diffraction image, Figure 4.18(c), that area contains noisy, chaotic energy, which based on its appearance in the corresponding gathers is not related to diffraction.

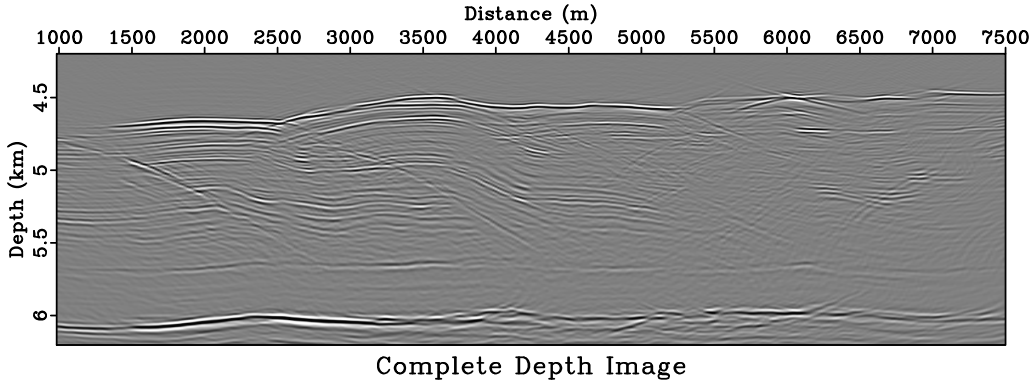
A final noteworthy feature is visible between 6.7 and 6.85 in three gathers. Here a polarity reversal in flat diffraction energy is visible for slope values near 0.001 s^2/m. This polarity reversal is an excellent example of diffractions caused by an edge rather than a point (Klem-Musatov et al., 2008). In this case, the diffraction is likely caused by a fault fault which has created a material discontinuity. This illustrates a situation where the assumption that diffractions are laterally coherent across all slopes, which is valid for point diffractions and justifies the use of gather semblance as a measure of diffraction, may not hold. Fortunately, in this case these edge diffractions

are successfully resolved by the probabilistic process, the corresponding event can be seen at 4100 m beginning around 6.7 s in Figure 4.18(d).

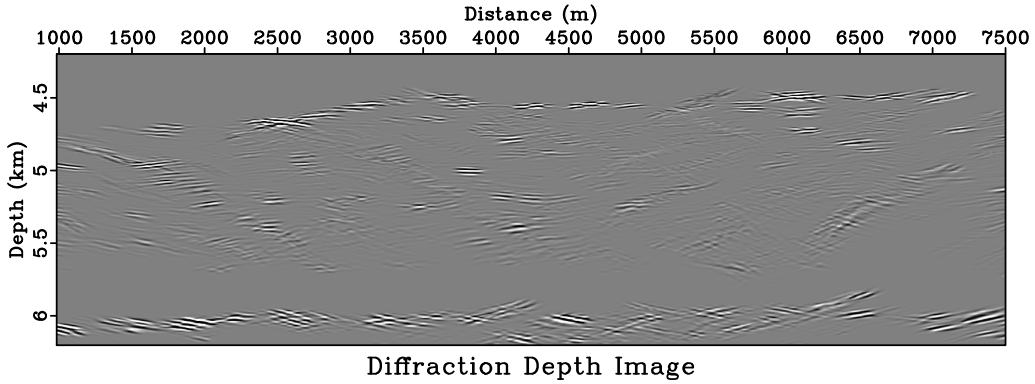
The probabilistic diffraction imaging process outputs a plausible RMS expectation velocity, Figure 4.17(a) using velocity analysis on data already stacked to zero offset. Because the calculated velocity field is relatively stable, I am able to calculate Dix velocity and transform our images from time to the depth domain in the manner of Sripanich and Fomel (2018).

Figures 4.20(b) and 4.20(a) contain the probabilistic diffraction image of Figure 4.18(d) and complete image of Figure 4.20(a) transformed to the depth domain and plotted with a true aspect ratio relating their horizontal and vertical components. Features, including the décollement near 5.6 km depth, the crystalline rock transition near 6 km depth, the seafloor, and thrust faults are visible in the diffraction and complete images. Reflections corresponding to highly deformed strata in an accretionary prism undergoing shortening and thickening and extending from the décollement to the seafloor are visible in the reflection image. In this image, the overriding Eurasian plate has relative motion to the left, towards the Philippine Sea and the subducting Philippine Plate has relative motion to the right, towards the Japanese island of Honshū.

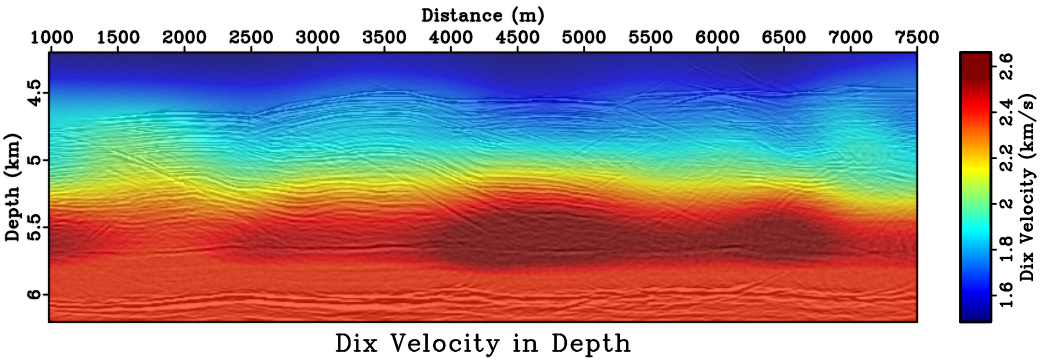
I calculate Dix, or interval, velocity using the RMS velocity field, and transform it to the depth domain. Depth domain Dix velocity is overlaid by the depth stretched complete image, and shown in Figure 4.20(c) to provide context for the velocity field. Velocities calculated in the probabilistic imaging process are reasonable when



(a)



(b)



(c)

Figure 4.20: Nankai Trough images and velocities transformed from the time to depth domains: (a) complete image; (b) probabilistic diffraction image; (c) Dix velocity overlaid with complete image.

compared to seismic velocity measurements in the region performed by the Integrated Ocean Drilling Program using core data and bore holes in the area (Moore et al., 2009). Examining the Dix velocity, a general trend of increasing velocity with depth is visible until the area of the décollement near 5.7 km depth, were velocity begins decreasing. Additionally, areas where the décollement is intersected by faults tend to have higher Dix velocities than surrounding areas. This is related to the fact that thrust faults extending from the décollement to the seafloor act as conduits facilitating the dewatering and compaction of sediments overlaying the décollement. The thrust faults do not extend below the décollement, so water has more difficulty escaping, hindering compaction and leading to lower velocities as well as a low velocity anomaly below the feature. This anomalous low corresponds to the deepest reliable velocity information from this study. Because I do not observe diffractions below the layer of strong diffractions at approximately 6 km marking the transition to crystalline rock, this study does not provide information about the presumably higher velocities underlaying that transition.

CONCLUSIONS

I formulate and apply a probabilistic approach to seismic diffraction imaging. By treating the weight functions in path-integral imaging as diffraction likelihood, I am able to emphasize wavefield components in seismic images output by OVC that are the most likely to correspond to a properly migrated seismic diffraction image and suppress the wavefield components that are not likely diffractions, improving the signal to noise ratio. The toy model example illustrated how the probabilistic imaging process

attenuated reflection energy and amplifies diffraction energy without explicit separation of the two, but rather as a result of applying weights to partial images. Thus, the method may be used in tandem with data-domain diffraction separation techniques to suppress any remnant reflection energy, as was done in the Nankai trough field data example. The process may attenuate some high frequency diffraction information, making it better suited for imaging diffractions in noisy environments. The synthetic experiment in this paper illustrated how the probabilistic method can be robust in such environments, creating an image featuring strong diffractions and suppressing noise with a RMS value 16 times greater than that of the noiseless diffraction energy. That synthetic experiment also demonstrated how the method has greater success imaging strong diffractions than weak ones and may have difficulty imaging diffractions whose moveout tails are superimposed by the tails of stronger overlaying diffractions. Diffractions imaged by the probabilistic process are laterally coherent in slope gathers because the weights are built from gather semblance. Correctly imaged diffraction energy is typically laterally coherent, but this assumption may be violated when energy from other diffractions becomes superimposed on a the slope gather centered above a correctly migrated diffraction. Therefore, this method should not be thought of as seeking to image every single diffractor within a seismic volume, or being superior to deterministic diffraction imaging, but rather as a supplementary tool to conventional diffraction imaging methods, outputting a image that identifies features that I can say with some certainty are correctly migrated diffractions. This can aid in the process of identifying geologically interesting features like the seafloor, faults, décollement or the transition from sedimentary to crystalline rock, seen in the

field data example from the Nankai trough.

Creating a probabilistic diffraction image through the proposed process does not require a migration velocity as an input, but rather generates one as an output. This means that direct comparison of the diffractions resolved by the deterministic image and the probabilistic image in the noiseless synthetic experiment as a measure of the probabilistic method's utility is not particularly meaningful, as that deterministic image in a noiseless environment with perfect apriori information of the subsurface velocity is effectively the best diffraction image one could hope to achieve in a single migration (although least-squares or sparse inversion techniques could yield a better resolved diffraction image iteratively (Merzlikin et al., 2020)). Instead, the probabilistic imaging method determines the most likely velocity field, as well as a measure of velocity uncertainty, as it operates. That output migration velocity may also be determined using a single offset of data, as was the case in the experiments presented in this paper. The expectation velocity produced in the synthetic experiments presented here tracked the correct migration velocity to within one standard deviation, and for the Nankai Trough field data example output reasonable velocities that resolved a velocity inversion. The process of finding the expectation velocity requires diffraction information, so the method's ability to determine correct migration velocities may be limited where diffractions are not present, as seen in the toy model example and in the Nankai Trough example below the transition to crystalline rock.

The challenges encountered in the studies featured in this paper introduce some promising directions for future inquiry. The difficulties faced by the probabilistic method at resolving a large dynamic range of diffractions could be mitigated by

decreasing the dynamic range of the weight functions by using, for example, powers of the combined weights, although that would also make the method more susceptible to noise. Finding the powers of weight functions that best balance highlighting diffraction signal and noise could improve the method. Similarly, finding more sophisticated ways of normalizing the weights to preserve weak diffraction and exploring other weights tied to the likelihood of diffraction could yield improved results. Particularly interesting is exploring different measures of diffraction "flatness", or diffraction likelihood, than semblance in the slope-gather domain. Such measures may not be ideal in the presence of edge diffractions, as seen in the field data example. Although the probabilistic imaging process using semblance was able to resolve those edge diffractions, a better method may exist. Applying the weighted imaging process to 3D Oriented Velocity Continuation should also enhance the output diffraction images, as that would be able to account for the possible effects of out of plane diffractions and reduce the effect of intersecting diffraction tails on semblance. The concept of probabilistic weighting for path-integral imaging can be used for collections of images where the wavefront is parameterized by more variables than just migration velocity, making application to sets images output by other time migration parameters with a continuation operators a fascinating direction of inquiry. As an anisotropic continuation operator already exists, pursuing a similar framework to determine weights based on both velocity and anisotropy, as well as using the method to determine conditional probability for anisotropy for each migration velocity, seems promising.

The ability of the method to suppress noise, highlighted in the noisy synthetic experiment, suggests that applying this method to passive seismic data could be

beneficial and makes it a promising direction of future inquiry. Geophone station signals could be used as input data for the continuation process with an extra event time, t_o variable added. This variable would correspond to earthquake event time in the seismic record. This extra variable could be applied as a shift on t prior to performing OVC and the probabilistic imaging process. Weights could be treated as conditional probabilities given event time t_o. Parallelization in t_o would be a relatively straightforward process, making the expensive required computations feasible.

Probabilistic path-integral diffraction imaging does not require advance knowledge of the migration velocity, which it generates as an output. Additional weights may be used to improve the results, and the method could be modified to emphasize different portions of the wavefield. Theoretically, this approach should function for not just velocity, but any time migration parameter with a continuation operator, so application to anisotropic imaging is a promising direction for future work.

Chapter 5

A variational approach for picking optimal surfaces from semblance-like panels using continuation

I propose and demonstrate a variational method for determining optimal velocity fields from semblance-like volumes using continuation. The proposed approach finds a minimal cost surface through such a volume, often a velocity field within a semblance scan, allowing picked velocity fields to incorporate information from gathers that are spatially near the midpoint in question. The minimization process amounts

[1]Some of the material in this chapter was published or submitted as:

- Decker, L., and S. Fomel, 2020, A variational method for picking velocity surfaces from semblance scans: SEG Technical Program Expanded Abstracts 2020, 3684–3688. The authors contributed to this paper as follows: study conception and design: S. Fomel; data collection: L. Decker; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.
- Decker, L., and S. Fomel, 2021a, A continuation approach for avoiding local minima in seismic velocity picking: SEG Technical Program Expanded Abstracts 2021, Submitted. The authors contributed to this paper as follows: study conception and design: L. Decker; data collection: L. Decker; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.
- Decker, L., and S. Fomel, 2021c, A variational approach for picking optimal surfaces from semblance-like panels using continuation: Geophysics, Submitted. The authors contributed to this paper as follows: study conception and design: L. Decker and S. Fomel; data collection: L. Decker; analysis and interpretation of results: L. Decker and S. Fomel; draft manuscript preparation: L. Decker. All authors reviewed results and approved the final version of the manuscript.

to solving a non-linear elliptic partial differential equation, which is accomplished by changing the problem to a parabolic problem and solving it iteratively until it converges to a critical point which minimizes the cost functional. The continuation approach functions by using a variational framework to iteratively minimize the cost of a velocity surface through successively less smoothed semblance scans, and works because a global minima for the velocity cost functional can only exist when the semblance scan varies smoothly in space and is convex in the parameter being scanned. I propose a velocity functional and prove both that it possesses a minimizer, and that a gradient descent scheme will converge to such a minimizer in an infinite dimensional setting, assuming that there exists a ball in H^1 with non-zero radius surrounding that minimizer. Using a discretization of the functional with a ℓ-BFGS algorithm I illustrate how the continuation approach is able to avoid local minima that capture the iterative solution of a velocity field without continuation and find a lower cost final model which can be used in seismic processing on a field dataset from the Viking Graben. I then employ a field dataset from the Gulf of Mexico to show how the final model output by the method when continuation is employed is largely independent of the starting velocity model, producing something resembling a global minimum. Finally, I illustrate the versatility of the variational picking approach by demonstrating how it may be used for automatic interpretation of a seismic horizon from the Heidrun field.

INTRODUCTION

The concept of picking normal moveout (NMO) velocity from the dominant semblance trend in velocity spectral display panels was pioneered by Taner and Koehler (1969). These panels measure how well applying a parameterized operation, like the NMO correction, over a parameter sweep transform data in some metric. For the case of Taner and Koehler (1969) this meant NMO gather flatness as measured by semblance, but numerous other geophysical applications exist including dip moveout (DMO) velocity analysis (Hale, 1984; Deregowski, 1986; Yilmaz, 2001), migration velocity analysis (Fowler, 1988; Deregowski, 1990; Fomel, 2003b; Decker et al., 2017a), image registration with different time shifts (Hale, 2013), data alignment by local similarity scan and similar applications (Fomel, 2007a, 2009b; Bader et al., 2019), deconvolution with dynamic frequency wavelets (Decker and Fomel, 2018), picking geobodies, faults, and seismic horizons in automatic interpretation (Wu and Fomel, 2018a,b; Yan and Wu, 2021), and determining the principal anisotropic axis in image gathers (Decker and Zhang, 2020).

Manually selecting such trends can be a labor intensive process, so a rich tradition of research has focused on increasing the ability of computers to learn the dominant trends in semblance-like panels with reduced need for human intervention. This has involved overcoming difficulties in the picking process where artificially or anomalously high semblance values do not reflect a high quality fit, so care must be taken to avoid outputting an unphysical velocity field. These artificially high values are often caused by artifacts related to lateral velocity variations (Hubral and Krey,

1980). Early efforts, like those of Sherwood and Poe (1972), involved interpolating a line between the largest semblance values detected, but this required interpreter supervision to avoid artifacts and unphysical velocity trends. de Bazelaire (1988) used the theory of geometric optics (Born and Wolf, 1959) to devise an "optical stack" method that automatically output velocity information. Doicin et al. (1994) modified the optical stack approach by applying constraints which force the method to avoid unphysical velocity selections. Alder and Brandwood (1999) expanded on the method of Doicin et al. (1994) by using three dimensional interpolation and locally scaled regression to determine a dense three dimensional velocity field. This work was further extended by Siliqi et al. (2003) to simultaneously solve for two parameters: velocity and an annellipticity term which can help account for the effects of dipping reflectors or seismic anisotropy. Arnaud et al. (2004) explored how these picking algorithms may be implemented in situations where the phase or amplitude of a seismic reflection changes with offset. Larner and Celis (2007) demonstrated how selective-correlation velocity analysis could be used to improve the resolution of velocity spectra and lead to more accurate picks. Research has continued on methods for automatic multiparameter scanning (Tao et al., 2012), including the efficient application of the three dimensional version of the Fourier integral operator butterfly algorithm (Candès et al., 2009) to the problem by Hu et al. (2015).

The approach outlined here follows a branch of inquiry began by Toldi (1989), who proposed a method for finding the best path through semblance iteratively given apriori knowledge of the likely velocity field and enforcing penalties for large changes in velocity. Symes and Carazzone (1991) proposed applying a variational principle in-

vert for reflection velocity models using differential semblance. Differential semblance is advantageous because in the case of noiseless data it possesses a single minima and exhibits convex behavior near that minimum, but with a reduction in velocity resolution compared to regular semblance (Symes, 1998, 1999; Mulder and ten Kroode, 2002; Li and Symes, 2007). Harlan (2001) simplified the method of Toldi (1989), replacing the apriori constraints of that method with a stiffness penalty and explicitly casting the ideal path through the semblance panel as the maximization of the variational integral:

$$\max_{v(\mathbf{x})} \int \alpha \left[v(\mathbf{x}), \mathbf{x} \right] d\mathbf{x}, \tag{5.1}$$

where $v(\mathbf{x})$ is a smooth surface defining the velocity and α is semblance. This maximization was performed using a Gauss-Newton algorithm (Luenberger and Ye, 1984). Although Harlan (2001) proposed a framework for solving for $v(\mathbf{x})$, a multidimensional surface, no examples were provided.

Inspired by the work of Deschamps and Cohen (2001) in virtual endoscopy, Fomel (2009b) continued with the variational approach of Harlan (2001). Fomel (2009b) noted that in the one dimensional case of a velocity path through a single gather, Equation 5.1 could be reformulated to appear analogous to a ray-tracing equation solving for the minimal travel time of the following integral,

$$\min_{v(t)} \int_{t_o}^{t_f} \exp\left(-\alpha \left[v(t), t \right]\right) \sqrt{\lambda^2 + \left(\frac{dv}{dt} \right)^2} dt, \tag{5.2}$$

where λ is a parameter that modifies the cost of changing position in t relative to changing position in v. Using variational methods (Lanczos, 1966; Greenberg, 1978;

Gelfand and Fomin, 2000), the optimal $v(t)$ in Equation 5.2 is determined by solving the eikonal equation (Babich and Buldyrev, 1972; Yilmaz, 2001) for $U(t, v)$,

$$\left(\frac{\partial}{\partial v}U(t,v)\right)^2 + \frac{1}{\lambda^2}\left(\frac{\partial}{\partial t}U(t,v)\right)^2 = \exp\left(-2\alpha\left[v(t),t\right]\right), \tag{5.3}$$

which, given a $v(t_o)$, may be done using a finite difference algorithm (Iserles, 1996). After determining $U(v,t)$, the method finds $v(t)$ by tracking backward along ∇U from the $v(t_f)$ that minimizes $U(v,t_f)$. Oscillations in $v(t)$ are dampened using shaping regularization (Fomel, 2007b).

I propose to expanded the approach of Fomel (2009b) to picking a multidimensional surface by minimizing a functional resembling semblance-weighted total variation regularization using a gradient descent method. This formulation enabled direct use of information from spatially adjacent semblance panels to determine a continuous velocity field without explicitly enforcing smoothing. This improves upon the existing one-dimensional, gather-by-gather approach of Fomel (2009b) by incorporating spatially adjacent information into the picking algorithm. This formulation is equivalent to a non-linear elliptic partial differential equation, which can be challenging to solve directly, so instead I propose to iteratively find minimizing surfaces iteratively. However, because an iterative method seeks out the nearest minimizer for the velocity functional, which can be highly multimodal, it may require an accurate starting model. Additionally, gradient descent methods, which may be used for iteratively finding a minima for a functional, may be slow to converge, requiring many iterations.

Continuation, or graduated optimization (Blake and Zisserman, 1987; Chapelle et al., 2006; Chaudhuri and Solar-Lezama, 2011; Mobahi and Fisher, 2015; Hazan et al., 2016; Xue et al., 2016), refers to a method of non-convex optimization where local minima may be avoided by solving a series of successively more challenging, or less convex, approximations to an optimization problem, and using the solution of a smoother problem as the starting model for more rugose one. This is frequently done by convolving the objective function with a Gaussian kernel (Wu, 1996). Gaussian convolution may be prohibitively expensive, but may be efficiently approximated by triangle smoothing (Claerbout, 1993).

The limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm (ℓ-BFGS) refers to a class of quasi-Newton schemes for accelerating the convergence of an iterative minimizer without the need for a large amount of computer memory (Nocedal, 1980; Liu and Nocedal, 1989; Li and Fukushima, 2001). ℓ-BFGS works by using several gradient computations to build an approximation of a system's Hessian. Such memory considerations become essential for large problems, as the Hessian is on the order of the number of samples in the input vector squared. Using the approximate Hessian enables the algorithm to draw on information about the curvature of cost function level sets, and thus provide a step direction leading more directly to a minimum.

In the following sections I propose an extension of Fomel (2009b) for picking velocity surfaces from semblance volumes and prove that the extension possesses a minimizer in H^1, and that a gradient descent scheme will converge to that minimizer assuming there exists a ball in H^1 with non-zero radius around that minimizer. The

problem is then discretized, and a ℓ-BFGS algorithm is used rather than gradient descent to accelerate convergence. I then demonstrate how continuation may be applied to the picking algorithm to bypass local minima and make the algorithm behave more like a global optimizer. Finally, the automatic picking algorithm is applied to an automatic interpretation problem to illustrate its versatility and applicability beyond picking velocity surfaces from semblance scans.

THEORY

Mathematical Treatment

As a natural extension of the one dimensional functional of Fomel (2009b) displayed in Equation 5.2 to higher dimensions, consider the functional $\tilde{F} : V \to \mathbb{R}^+$, defined

$$\tilde{F}[v] = \int_\Omega e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \sqrt{\lambda^2 + |\nabla v|^2} d\Omega. \tag{5.4}$$

In Equation 5.4 I assume that $0 < \lambda \in \mathbb{R}$, $\Omega \subset \mathbb{R}^n$ is a convex bounded set with Lipschitz boundary $\partial\Omega = \Gamma$, real valued functions $V \subset H^1(\Omega)$ such that $V \ni v \cdot \hat{\mathbf{n}} = 0$ for $\hat{\mathbf{n}} \in \Gamma$, $\alpha \in C^1(V \times \Omega, [0, \bar{\alpha}])$ with $0 < \bar{\alpha}$. $\alpha[v, \mathbf{x}]$ is assumed to be a Lipschitz continuous function in v such that $|\alpha_v| \leq L_{\alpha 1}$, and whose partial derivative in v is also Lipschitz continuous with $|\alpha_{vv}| \leq L_{\alpha 2}$. I use $|x|$ to signify $\sqrt{x \cdot x}$ for $x \in \mathbb{R}^n$, which reduces to the absolute value function if $x \in \mathbb{R}$. I wish to find the $v \in V$ that minimizes this equation.

I define $F(\mathbf{x}, v, p)$ to be the integrand of Equation 5.4,

$$F(\mathbf{x}, v, p) = e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \sqrt{\lambda^2 + |p|^2}, \tag{5.5}$$

and compute its first order partial derivatives

$$F_v(\mathbf{x}, v, p) = -\alpha_v[v(\mathbf{x}), \mathbf{x}]e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}\sqrt{\lambda^2 + |p|^2}, \tag{5.6}$$

and

$$F_p(\mathbf{x}, v, p) = \frac{e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}}{\sqrt{\lambda^2 + |p|^2}}p. \tag{5.7}$$

Based on the limits placed on $\alpha \in [0, \tilde{\alpha}]$ and $\left|\frac{\partial}{\partial v}\alpha\right| = |\alpha_v| \leq L_{\alpha 1}$ I can bound these derivatives as

$$|F_v(\mathbf{x}, v, p)| \leq L_{\alpha 1}\left(\lambda + |p|\right), \tag{5.8}$$

and

$$|F_p| \leq \sqrt{2}, \tag{5.9}$$

for which I use the inequality $\sqrt{a^2 + b^2} \geq \frac{1}{\sqrt{2}}\left(|a| + |b|\right)$ for $a, b \in \mathbb{R}$. To see how this is true, consider the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ defined $\mathbf{x} = [|a|, |b|]^T$, and $\mathbf{y} = [1, 1]^T$. Using the Cauchy-Schwartz inequality

$$|\langle \mathbf{x}, \mathbf{y}\rangle_{\ell^2}|^2 \leq ||\mathbf{x}||^2_{\ell^2}\,||\mathbf{y}||^2_{\ell^2}\,. \tag{5.10}$$

Because $\langle \mathbf{x}, \mathbf{y}\rangle_{\ell^2} = |a| + |b|$, $||\mathbf{x}||^2_{\ell^2} = a^2 + b^2 + 2\,|a|\,|b|$, and $||\mathbf{y}||^2_{\ell^2} = 2$,

$$\left(|a| + |b|\right)^2 \leq 2\left(a^2 + b^2 + 2\,|a|\,|b|\right), \tag{5.11}$$

and noting that $|a|, |b| \geq 0$,

$$\frac{|a| + |b|}{\sqrt{2}} \leq \sqrt{a^2 + b^2}. \tag{5.12}$$

I can also compute the second order partial derivatives of F,

$$F_{vv}(\mathbf{x}, v, p) = \left(\alpha_v^2[v(\mathbf{x}), \mathbf{x}] - \alpha_{vv}[v(\mathbf{x}), \mathbf{x}]\right)e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}\sqrt{\lambda^2 + |p|^2}, \tag{5.13}$$

$$F_{vp}(\mathbf{x}, v, p) = -\alpha_v[v(\mathbf{x}), \mathbf{x}] \frac{e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}}{\sqrt{\lambda^2 + |p|^2}} p, \tag{5.14}$$

and

$$F_{pp}(\mathbf{x}, v, p) = \frac{\lambda^2 e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}}{\left(\lambda^2 + |p|^2\right)^{\frac{3}{2}}}. \tag{5.15}$$

Similarly, I may bound these derivatives using

$$|F_{vv}(\mathbf{x}, v, p)| \leq \left(L_{\alpha 1}^2 + L_{\alpha 2}\right)(\lambda + |p|), \tag{5.16}$$

$$|F_{vp}(\mathbf{x}, v, p)| \leq \sqrt{2} L_{\alpha 1}, \tag{5.17}$$

and

$$0 < F_{pp} \leq \frac{1}{\lambda}. \tag{5.18}$$

Note that although $0 < F_{pp}$, in the limit where $|p| \to \infty$, $F_{pp} \to 0$. Because these partial derivatives of F exist everywhere $\tilde{F}(v)$, $F(\mathbf{x}, v, p)$, $F_v(\mathbf{x}, v, p)$, and $F_p(\mathbf{x}, v, p)$ are continuous. The continuity of these multivariable functions flows from the fact that for a function of two variables A and appropriate norm $||\cdot||$,

$$\begin{aligned} ||A(x_1, y_1) - A(x_2, y_2)|| &= ||A(x_1, y_1) - A(x_2, y_1) + A(x_2, y_1) - A(x_2, y_2)|| \\ &\leq ||A(x_1, y_1) - A(x_2, y_1)|| + ||A(x_2, y_1) - A(x_2, y_2)||, \end{aligned} \tag{5.19}$$

using the triangle inequality. Subsequently

$$\begin{aligned} \lim_{x,y \to x_0,y_0} ||A(x, y) - A(x_0, y_0)|| \leq \lim_{x,y \to x_0,y_0} &||A(x, y) - A(x_0, y)|| \\ &+ ||A(x_0, y) - A(x_0, y_0)||. \end{aligned} \tag{5.20}$$

Therefore, if A is continuous in both x and y it is continuous. Several of the partial derivatives above are bound with $|p| = |\nabla v|$, so they are not globally Lipschitz continuous. To aid later analysis of this concept, I will use the introduce the concept of local Lipschitz continuity.

Definition 5.1 (Locally Lipschitz Continuous)**.** *Given two Banach spaces, U and W, an operator $\mathcal{F} : U \to W$ is called* locally Lipschitz continuous *if for every bounded subset $S \subset U$ the restriction of the operator $\mathcal{F}|_S$ is Lipschitz continuous. Therefore there exists a $L_S = L(\mathcal{F}, S)$ such that*

$$||\mathcal{F}(x) - \mathcal{F}(y)||_W \leq L_S \, ||x - y||_U \;\; \forall x, y \in S. \tag{5.21}$$

The argument used in Equation 5.20 may be applied to Lipschitz (and local Lipschitz) continuity for functions of multiple variables. Therefore, if a multivariable function is (locally) Lipschitz in each argument it is (locally) Lipschtiz.

I seek to minimize the functional in Equation 5.4. To do this in a finite dimensional setting, one would take the derivative of that functional with respect to its arguments and determine the critical points where the first derivative is equal to zero. Minima then occur at critical points where the second derivative or Hessian is positive. To facilitate my search for a minima, I will introduce two concepts of a derivative in this infinite dimensional setting, the Gateaux derivative and the Fréchet derivative.

Definition 5.2 (Gateaux Derivative)**.** *Suppose U and W are two locally convex topological vector spaces (Banach spaces fulfill this criteria) with $X \subset U$ open and $\mathcal{F} : U \to W$. The Gateaux Derivative $d\mathcal{F}(u; \psi)$ of $\mathcal{F}$ at $u \in U$ in direction $\psi \in X$ is defined as*

$$\begin{aligned} d\mathcal{F}(u; \psi) &= \lim_{\rho \to 0} \frac{\mathcal{F}(u + \rho\psi) - \mathcal{F}(u)}{\rho} \\ &= \frac{d}{d\rho} \mathcal{F}(u + \rho\psi) \Big|_{\rho=0}. \end{aligned} \tag{5.22}$$

This is a generalization of the directional derivative. When $d\mathcal{F}(u,\psi)$ exists for all $\psi \in X$, $\mathcal{F}$ is said to be Gateaux differentiable *at u.*

Definition 5.3 (Fréchet Derivative)**.** *Suppose U and W are normed vector spaces (Banach spaces are complete normed vector spaces). Let $X \subset U$ be open. A function $\mathcal{F} : X \to W$ is said to be* Fréchet differentiable *at $x \in X$ if there exists a bounded linear operator $\mathcal{A} : U \to V$ such that for $\xi \in U$*

$$\lim_{\xi \to 0} \frac{||\mathcal{F}(u+\xi) - \mathcal{F}(u) - \mathcal{A}\xi||_W}{||\xi||_U} = 0, \tag{5.23}$$

when such an operator exists, $D\mathcal{F}(u) = \mathcal{A}$ is called the Fréchet derivative. If such an operator exists it coincides with the Gateaux derivative, but a function may be Gateaux differentiable without being Fréchet differentiable. A functional $f : X \to \mathbb{R}$ is said to be a member of $C^1(X, \mathbb{R})$ if its Fréchet derivative exists and is continuous.

Recall, $V \subset H^1(\Omega)$, a Hilbert space (complete inner product space) with the inner product

$$\begin{aligned} \langle x, y \rangle_{H^1(\Omega)} &= \langle x, y \rangle_{L^2(\Omega)} + \langle \nabla x, \nabla y \rangle_{L^2(\Omega)} \\ &= \int_\Omega xy \, d\Omega + \int_\Omega \nabla x \cdot \nabla y d\Omega, \end{aligned} \tag{5.24}$$

for $x, y \in H^1(\Omega)$, and induced norm

$$\begin{aligned} ||x||^2_{H^1(\Omega)} &= ||x||^2_{L^2(\Omega)} + ||\nabla x||^2_{L^2(\Omega)} \\ &= \int_\Omega |x|^2 \, d\Omega + \int_\Omega |\nabla x|^2 \, d\Omega. \end{aligned} \tag{5.25}$$

This norm enables characterization of the functions in $H^1(\Omega)$ for the use of this analysis:

$$H^1(\Omega) = \{f : \Omega \to \mathbb{R} \mid \ ||f||_{H^1(\Omega)} < \infty\}. \tag{5.26}$$

I now determine the Gateaux derivative of Equation 5.4 and show that it coincides with the Fréchet derivative.

Proposition 5.1. *The functional in Equation 5.4 is Fréchet differentiable, with Fréchet derivative*

$$D\tilde{F}(u)(\psi) = \langle \psi, F_v(\mathbf{x}, u, \nabla u) \rangle_{L^2(\Omega)} + \langle \nabla\psi, F_p(\mathbf{x}, u, \nabla u) \rangle_{L^2(\Omega)} . \tag{5.27}$$

Proof. Applying the Gateaux derivative to Equation 5.4 provides

$$d\tilde{F}(v;\psi) = \left. \frac{d}{d\rho} \tilde{F}(v + \rho\psi) \right|_{\rho=0} . \tag{5.28}$$

Because $\tilde{F}(v)$ and $F(\mathbf{x}, v, p)$ are continuous functions, this differential may be written as

$$d\tilde{F}(v;\psi) = \int_\Omega \left. \frac{d}{d\rho} F(\mathbf{x}, v + \rho\psi, \nabla v + \rho\nabla\psi) \right|_{\rho=0} d\Omega, \tag{5.29}$$

which is equivalent to

$$\begin{aligned} d\tilde{F}(v;\psi) = &\int_\Omega \psi F_v(\mathbf{x}, v + \rho\psi, \nabla v + \rho\nabla\psi)|_{\rho=0} \, d\Omega \\ &+ \int_\Omega \nabla\psi \cdot F_p(\mathbf{x}, v + \rho\psi, \nabla v + \rho\nabla\psi)|_{\rho=0} \, d\Omega. \end{aligned} \tag{5.30}$$

Evaluating at $\rho = 0$ provides the Gateaux derivative,

$$\begin{aligned} d\tilde{F}(v;\psi) = &\int_\Omega \psi F_v(\mathbf{x}, v, \nabla v) d\Omega \\ &+ \int_\Omega \nabla\psi \cdot F_p(\mathbf{x}, v, \nabla v) d\Omega. \end{aligned} \tag{5.31}$$

Equation 5.31 may be written as a continuous linear operator where $(\cdot, \cdot)$ is the duality $(\mathcal{H}^*, \mathcal{H})$ for Hilbert space $\mathcal{H}$ and its dual $\mathcal{H}^*$

$$d\tilde{F}(v;\psi) = \left(d\tilde{F}(v), \psi \right) = \langle F_v(\mathbf{x}, v, \nabla v), \psi \rangle_{L^2(\Omega)} + \langle F_p(\mathbf{x}, v, \nabla v), \nabla\psi \rangle_{L^2(\Omega)} . \tag{5.32}$$

Substituting in Equations 5.6 and 5.7 provides

$$\begin{aligned} d\tilde{F}(v;\psi) = &- \int_{\Omega} \psi \alpha_v[v(\mathbf{x}),\mathbf{x}] e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \sqrt{\lambda^2 + |\nabla v|^2} d\Omega \\ &+ \int_{\Omega} \nabla\psi \cdot \nabla v \frac{e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}}{\sqrt{\lambda^2 + |\nabla v|^2}} d\Omega. \end{aligned} \tag{5.33}$$

Equation 5.33 is defined for all $\psi \in H^1(\Omega)$ and $v \in V$, so Equation 5.4 is Gateaux differentiable. I now show that the Gateaux derivative is also the Fréchet derivative.

To do this I wish to show that

$$\lim_{||\xi||_{H^1(\Omega)} \to 0} \frac{\left| \tilde{F}(u+\xi) - \tilde{F}(u) - d\tilde{F}(u,\xi) \right|}{||\xi||_{H^1(\Omega)}} = 0. \tag{5.34}$$

Let

$$\Psi(u,\xi) = \frac{\left| \tilde{F}(u+\xi) - \tilde{F}(u) - d\tilde{F}(u,\xi) \right|}{||\xi||_{H^1(\Omega)}}, \tag{5.35}$$

which may be rewritten as

$$\Psi(u,\xi) = ||\xi||^{-1}_{H^1(\Omega)} \left| \int_{\Omega} \left[F(\mathbf{x}, u+\xi, \nabla u + \nabla\xi) - F(\mathbf{x}, u, \nabla u) \right] d\Omega - d\tilde{F}(u,\xi) \right|. \tag{5.36}$$

Because the output of a norm is always greater than or equal to zero,

$$\Psi(u,\xi) \geq 0. \tag{5.37}$$

Using the argument shown in Equation 5.19 and letting $v = u + \xi$,

$$\begin{aligned} |F(\mathbf{x}, v, \nabla v) - F(\mathbf{x}, u, \nabla u)| \leq &|F(\mathbf{x}, v, \nabla v) - F(\mathbf{x}, u, \nabla v)| \\ &+ |F(\mathbf{x}, u, \nabla v) - F(\mathbf{x}, u, \nabla u)|. \end{aligned} \tag{5.38}$$

Recall that $F(\mathbf{x}, v, p)$ is continuous in v and p, so I may apply the mean value theorem

$$\begin{aligned} |F(\mathbf{x}, v, \nabla v) - F(\mathbf{x}, u, \nabla u)| \leq & \sup_{t\in(0,1)} |F_v(\mathbf{x}, u + t(v-u), \nabla v)| \, |v-u| \\ & + \sup_{t\in(0,1)} |F_p(\mathbf{x}, u, \nabla u + t(\nabla v - \nabla u))| \, |\nabla v - \nabla u| \,. \end{aligned} \tag{5.39}$$

Thus, I define the functions

$$\gamma_v(w, \xi) = \sup_{t\in(0,1)} |F_v(\mathbf{x}, w + t\xi, \nabla w)| \,, \tag{5.40}$$

and

$$\gamma_p(w, \xi) = \sup_{t\in(0,1)} |F_p(\mathbf{x}, w, \nabla(w + t\xi))| \,, \tag{5.41}$$

which are finite for $\xi \in H^1(\Omega)$, $w \in V$. Additionally,

$$\lim_{||\xi||_{H^1(\Omega)}\to 0} \gamma_v(w, \xi) = |F_v(\mathbf{x}, w, \nabla w)| \,, \tag{5.42}$$

and

$$\lim_{||\xi||_{H^1(\Omega)}\to 0} \gamma_p(w, \xi) = |F_p(\mathbf{x}, w, \nabla w)| \,, \tag{5.43}$$

This enables me to write

$$\begin{aligned} |F(\mathbf{x}, v, \nabla v) - F(\mathbf{x}, u, \nabla u)| \leq & \gamma_v(u, v-u) \, |v-u| \\ & + \gamma_p(u, v-u) \, |\nabla v - \nabla u| \,. \end{aligned} \tag{5.44}$$

Substituting Equation 5.36,

$$\Psi(u, \xi) \leq ||\xi||^{-1}_{H^1(\Omega)} \left| \int_\Omega \left[\gamma_v(u, \xi) \, |\xi| + \gamma_p(u, \xi) \, |\nabla \xi| \right] d\Omega - d\tilde{F}(u, \xi) \right| , \tag{5.45}$$

and using Equation 5.31,

$$\begin{aligned} \Psi(u, \xi) \leq ||\xi||^{-1}_{H^1(\Omega)} \int_\Omega & |[\gamma_v(u, \xi) \, |\xi| - F_v(\mathbf{x}, u, \nabla u)\xi \\ & + \gamma_p(u, \xi) \, |\nabla \xi| - F_p(\mathbf{x}, u, \nabla u) \cdot \nabla \xi]| \, d\Omega. \end{aligned} \tag{5.46}$$

Because $\gamma_v(u,\xi) \geq F_v(\mathbf{x},u,\nabla u)$ and $\gamma_p(u,\xi) \geq F_p(\mathbf{x},u,\nabla u)$,

$$\begin{aligned}\Psi(u,\xi) \leq ||\xi||^{-1}_{H^1(\Omega)} \int_\Omega &\left[\left(\gamma_v(u,\xi) - |F_v(\mathbf{x},u,\nabla u)|\right)|\xi| \right. \\ &\left. + \left(\gamma_p(u,\xi) - |F_p(\mathbf{x},u,\nabla u)|\right)|\nabla\xi|\right] d\Omega.\end{aligned} \tag{5.47}$$

For ease of notation I introduce

$$A(u,\xi) = \gamma_v(u,\xi) - |F_v(\mathbf{x},u,\nabla u)|\,, \tag{5.48}$$

and

$$B(u,\xi) = \gamma_p(u,\xi) - |F_p(\mathbf{x},u,\nabla u)|\,, \tag{5.49}$$

for which $A, B \geq 0$. Additionally, Equations 5.42 and 5.43 show that

$$\lim_{||\xi||_{H^1(\Omega)}\to 0} A(u,\xi) = 0, \tag{5.50}$$

and

$$\lim_{||\xi||_{H^1(\Omega)}\to 0} B(u,\xi) = 0. \tag{5.51}$$

Using A and B I write

$$\Psi(u,\xi) \leq ||\xi||^{-1}_{H^1(\Omega)} \int_\Omega \left(A(u,\xi) + B(u,\xi)\right)\left(|\xi| + |\nabla\xi|\right) d\Omega, \tag{5.52}$$

which is the inner product

$$\Psi(u,\xi) \leq ||\xi||^{-1}_{H^1(\Omega)} \left\langle A(u,\xi) + B(u,\xi), |\xi| + |\nabla\xi| \right\rangle_{L^2(\Omega)}. \tag{5.53}$$

Using the Cauchy-Schwartz inequality and the triangle inequality,

$$\Psi(u,\xi) \leq ||\xi||^{-1}_{H^1} \left(||\xi||_{L^2(\Omega)} + ||\nabla\xi||_{L^2(\Omega)}\right)\left(||A(u,\xi)||_{L^2(\Omega)} + ||B(u,\xi)||_{L^2(\Omega)}\right). \tag{5.54}$$

Since $\sqrt{a^2+b^2} \geq \frac{1}{\sqrt{2}}(|a|+|b|)$ as shown in Equation 5.12 and using the definition of the $H^1(\Omega)$ norm,

$$\Psi(u,\xi) \leq \sqrt{2}\,||\xi||_{H^1}^{-1}\,||\xi||_{H^1(\Omega)}\left(||A(u,\xi)||_{L^2(\Omega)} + ||B(u,\xi)||_{L^2(\Omega)}\right). \tag{5.55}$$

Finally, using Equations 5.50 and 5.51,

$$\lim_{||\xi||_{H^1(\Omega)}\to 0} \Psi(u,\xi) \leq 0. \tag{5.56}$$

Combining Equations 5.37 and 5.56, as well as the definition of Ψ in Equation 5.35 I arrive at the desired result,

$$\lim_{||\xi||_{H^1(\Omega)}\to 0} \frac{\left|\tilde{F}(u+\xi) - \tilde{F}(u) - d\tilde{F}(u,\xi)\right|}{||\xi||_{H^1(\Omega)}} = 0, \tag{5.57}$$

and thus I may state the Fréchet derivative of Equation 5.4,

$$D\tilde{F}(u)(\psi) = \langle \psi, F_v(\mathbf{x},u,\nabla u)\rangle_{L^2(\Omega)} + \langle \nabla\psi, F_p(\mathbf{x},u,\nabla u)\rangle_{L^2(\Omega)}. \tag{5.58}$$

□

Proposition 5.2. *The Fréchet derivative of Equation 5.4 given by Equation 5.58 is locally Lipschitz continuous.*

Proof. Consider a bounded subset $S \subset B_{H^1(\Omega)}(0,R) \cap V = \{v \in V \mid ||v||_{H^1(\Omega)} < R\}$, the H^1 ball of radius R centered at $0_{H^1(\Omega)}$. For any $v \in S$, $||\nabla v||_{L^2(\Omega)} < R$ because $||v||^2_{H^1(\Omega)} = ||v||^2_{L^2(\Omega)} + ||\nabla v||^2_{L^2(\Omega)}$. I now compute how the Fréchet derivative changes between inputs $u, w \in S$

$$\begin{aligned}\left|D\tilde{F}(u)(\psi) - D\tilde{F}(w)(\psi)\right| = \Big|&\langle \psi, F_v(\mathbf{x},u,\nabla u) - F_v(\mathbf{x},w,\nabla w)\rangle_{L^2(\Omega)} \\ &+ \langle \nabla\psi, F_p(\mathbf{x},u,\nabla u) - F_p(\mathbf{x},w,\nabla w)\rangle_{L^2(\Omega)}\Big|.\end{aligned} \tag{5.59}$$

Utilizing the triangle inequality and Cauchy-Schwartz I may write

$$\begin{aligned}\left|D\tilde{F}(u)(\psi) - D\tilde{F}(w)(\psi)\right| \leq ||\psi||_{L^2} \, ||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \\ + ||\nabla\psi||_{L^2} \, ||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \, .\end{aligned} \tag{5.60}$$

Recalling Equation 5.19

$$\begin{aligned}||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq ||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla u)||_{L^2(\Omega)} \\ + ||F_v(\mathbf{x}, w, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \, .\end{aligned} \tag{5.61}$$

The bounds on the partial derivatives of F_v in Equation 5.16 and 5.17 can be applied for some $q \in B_{H^1(\Omega)}(0, R)$

$$\begin{aligned}||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq \left|\left|\left(L_{\alpha 1}^2 + L_{\alpha 2}\right)(\lambda + |\nabla q|)\,|u - w|\right|\right|_{L^2(\Omega)} \\ + \left|\left|\sqrt{2}L_{\alpha 1}\,|\nabla u - \nabla w|\right|\right|_{L^2(\Omega)} \, ,\end{aligned} \tag{5.62}$$

and using Cauchy-Schwartz and the absolute homogeneity of norms provides

$$\begin{aligned}&||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq \\ &\qquad \left(L_{\alpha 1}^2 + L_{\alpha 2}\right)(\lambda + ||\nabla q||_{L^2})\,||u - w||_{L^2(\Omega)} \\ &\qquad + \sqrt{2}L_{\alpha 1}\,||\nabla u - \nabla w||_{L^2(\Omega)} \, .\end{aligned} \tag{5.63}$$

Because $||\nabla q||_{L^2(\Omega)} < R$ I can write

$$\begin{aligned}||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq \left(L_{\alpha 1}^2 + L_{\alpha 2}\right)(\lambda + R)\,||u - w||_{L^2(\Omega)} \\ + \sqrt{2}L_{\alpha 1}\,||\nabla u - \nabla w||_{L^2(\Omega)} \, ,\end{aligned} \tag{5.64}$$

which using Equation 5.12 and the definition of the $H^1(\Omega)$ norm may again be bounded as

$$\begin{aligned}&||F_v(\mathbf{x}, u, \nabla u) - F_v(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq \\ &\qquad \sqrt{2}\left[\left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\,(\lambda + R) + \sqrt{2}L_{\alpha 1}\right]||u - w||_{H^1(\Omega)} \, .\end{aligned} \tag{5.65}$$

I now apply a similar argument to F_p

$$\begin{aligned} ||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq & ||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla u)||_{L^2(\Omega)} \\ & + ||F_p(\mathbf{x}, w, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} . \end{aligned} \tag{5.66}$$

Again proceeding in the manner of Equation 5.19

$$\begin{aligned} ||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq & ||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla u)||_{L^2(\Omega)} \\ & + ||F_v(\mathbf{x}, w, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} . \end{aligned} \tag{5.67}$$

Substituting the bounds in Equations 5.17 and 5.16 provides

$$\begin{aligned} ||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq & \sqrt{2} L_{\alpha 1} ||u - w||_{L^2(\Omega)} \\ & + \frac{1}{\lambda} ||\nabla u - \nabla w||_{L^2(\Omega)} , \end{aligned} \tag{5.68}$$

which utilizing Equation 5.12 may be bounded with

$$||F_p(\mathbf{x}, u, \nabla u) - F_p(\mathbf{x}, w, \nabla w)||_{L^2(\Omega)} \leq \left(2L_{\alpha 1} + \frac{\sqrt{2}}{\lambda} \right) ||u - w||_{H^1(\Omega)} . \tag{5.69}$$

I substitute Equations 5.65 and 5.69 into Equation 5.60

$$\begin{aligned} \left| D\tilde{F}(u)(\psi) - D\tilde{F}(w)(\psi) \right| \leq & \\ & \sqrt{2} \left[\left(L_{\alpha 1}^2 + L_{\alpha 2} \right) (\lambda + R) + \sqrt{2} L_{\alpha 1} \right] ||\psi||_{L^2(\Omega)} ||u - w||_{H^1(\Omega)} \\ & + \left(2L_{\alpha 1} + \frac{\sqrt{2}}{\lambda} \right) ||\nabla \psi||_{L^2(\Omega)} ||u - w||_{H^1(\Omega)} . \end{aligned} \tag{5.70}$$

Introducing a bounding constant $L_{D\tilde{F}}$

$$L_{D\tilde{F}} = 2 \left[\left(L_{\alpha 1}^2 + L_{\alpha 2} \right) (\lambda + R) + \sqrt{2} L_{\alpha 1} + \frac{1}{\lambda} \right] , \tag{5.71}$$

allows me to write

$$\left|D\tilde{F}(u)(\psi) - D\tilde{F}(w)(\psi)\right| \leq \frac{L_{D\tilde{F}}}{\sqrt{2}} ||u-w||_{H^1(\Omega)} \left(||\psi||_{L^2(\Omega)} + ||\nabla\psi||_{L^2(\Omega)}\right). \quad (5.72)$$

Finally, applying Equation 5.12

$$\left|D\tilde{F}(u)(\psi) - D\tilde{F}(w)(\psi)\right| \leq L_{D\tilde{F}} ||u-w||_{H^1(\Omega)} ||\psi||_{H^1(\Omega)}. \quad (5.73)$$

Therefore the Fréchet derivative of Equation 5.4 is locally Lipschitz continuous, with a local Lipschitz constant given by Equation 5.71. □

Having established the derivatives of $\tilde{F}$, I wish to find the analogue for a gradient of $\tilde{F}$. Using the Reisz representation theorem, this is the unique element in $H^1(\Omega)$, $\nabla\tilde{F}[u]$, such that $\left\langle \nabla\tilde{F}[u], w \right\rangle = \left(D\tilde{F}(u), (w)\right)$ for all u and $w \in H^1(\Omega)$. As Smyrlis and Zisis (2004) note, in general the gradient of functional $f : \mathcal{H} \to \mathbb{R}$ labeled $\nabla f \in \mathcal{H}$ for real Hilbert space $\mathcal{H}$ and its corresponding Fréchet derivative, $Df \in \mathcal{H}^*$ may also be characterized by $||\nabla f(w)||_{\mathcal{H}} = ||Df(w)||_{\mathcal{H}^*}$, and $(\nabla f(w), Df(w)) = ||Df(w)||^2_{\mathcal{H}^*}$ $\forall u \in \mathcal{H}$. Additionally, $||\nabla f(u) - \nabla f(w)||_{\mathcal{H}} = ||Df(u) - Df(w)||_{\mathcal{H}^*}$ $\forall u, w \in \mathcal{H}$. A convenient property of $H^1(\Omega)$ is that because it is based on L^2, and the dual of an L^p space is L^q for $\frac{1}{p} + \frac{1}{q} = 1$, $p, q \in \mathbb{N}$, $H^1(\Omega)^* = H^1(\Omega)$.

Using the divergence theorem,

$$\begin{aligned} \langle \nabla w, F_p(\mathbf{x}, u, \nabla u)\rangle_{L^2(\Omega)} = &- \langle w, \nabla \cdot F_p(\mathbf{x}, u, \nabla u)\rangle_{L^2(\Omega)} \\ &+ \int_\Gamma w F_p(\mathbf{x}, u, \nabla u) \cdot \hat{\mathbf{n}} d\Gamma. \end{aligned} \quad (5.74)$$

Substituting Equation 5.7

$$\int_{\Gamma} wF_p(\mathbf{x}, u, \nabla u) \cdot \hat{\mathbf{n}} d\Gamma = \int_{\Gamma} \frac{e^{-\alpha[u(\mathbf{x}),\mathbf{x}]}}{\sqrt{\lambda^2 + |\nabla u|^2}} \nabla u \cdot \hat{\mathbf{n}} d\Gamma, \tag{5.75}$$

and $\nabla u \cdot \hat{\mathbf{n}} = 0 \; \forall u \in V$, $\hat{\mathbf{n}} \in \Gamma$, so

$$D\tilde{F}(u)(w) = \langle w, F_v(\mathbf{x}, u, \nabla u) - \nabla \cdot F_p(\mathbf{x}, u, \nabla u) \rangle_{L^2(\Omega)}. \tag{5.76}$$

Because Equation 5.77 holds for all w, I can write the strong form of the gradient

$$\nabla \tilde{F}(u) = F_v(\mathbf{x}, u, \nabla u) - \nabla \cdot F_p(\mathbf{x}, u, \nabla u). \tag{5.77}$$

Substituting Equations 5.6 and 5.7 provides

$$\nabla \tilde{F}(u) = -\alpha_v[u(\mathbf{x}), \mathbf{x}] e^{-\alpha[u(\mathbf{x}),\mathbf{x}]} \sqrt{\lambda^2 + |\nabla u|^2} - \nabla \cdot \left(\frac{e^{-\alpha[u(\mathbf{x}),\mathbf{x}]}}{\sqrt{\lambda^2 + |\nabla u|^2}} \nabla u \right). \tag{5.78}$$

A small perturbation of u in the "direction" of $-\nabla \tilde{F}(u)$ will offer the greatest reduction in cost relative to a perturbation any other "direction" with the same magnitude. To see this, I can apply the Cauchy-Schwartz inequality

$$\frac{\left| \left\langle \nabla \tilde{F}(u), w \right\rangle_{L^2(\Omega)} \right|}{\left|\left| \nabla \tilde{F}(u) \right|\right|_{L^2(\Omega)} ||w||_{L^2(\Omega)}} \leq 1. \tag{5.79}$$

The left hand side achieves the maximum value of 1 when $w = \nabla \tilde{F}(u)$ assuming $\left|\left| \nabla \tilde{F}(u) \right|\right|_{L^2(\Omega)} = ||w||_{L^2(\Omega)}$. Additionally, in order for a function u to minimize Equation 5.4 it must be a critical point which fulfills $\nabla \tilde{F}(u) = 0$, or equivalently

$$\nabla \cdot \left(\frac{e^{-\alpha[u(\mathbf{x}),\mathbf{x}]}}{\sqrt{\lambda^2 + |\nabla u|^2}} \nabla u \right) = -\alpha_v[u(\mathbf{x}),\mathbf{x}] e^{-\alpha[u(\mathbf{x}),\mathbf{x}]} \sqrt{\lambda^2 + |\nabla u|^2}. \tag{5.80}$$

Equation 5.80 is a non-linear elliptic partial differential equation, which is challenging to solve directly. Instead, I choose to adopt an iterative approach, and instead solve the parabolic equation

$$\frac{\partial u}{\partial t} = -\nabla \tilde{F}(u), \tag{5.81}$$

until it converges to a critical u^* where $\left|\left|\nabla \tilde{F}(u^*)\right|\right| = 0$. This is a common approach for finding the minima of a functional (Mawhin and Willem, 2010), and suggests the use of an iterative gradient descent scheme. This method provides a sequence $\{u_i\}_{i\in\mathbb{N}}$ of subsequent approximations for for u^*, the critical point of $\tilde{F}(u)$. The update structure given an initial u_0 is

$$u_{i+1} = u_i + \rho_i h_i. \tag{5.82}$$

Control parameters, also known as step sizes, $\{\rho_i\}_{i\in\mathbb{N}}$ satisfy $a \leq \rho_i \leq b$ for positive real numbers a, b. At each step in Equation 5.82 $\nabla \tilde{F}(u_i)$ is evaluated. If $\left|\left|\nabla \tilde{F}(u_i)\right|\right| = 0$ the method stops because it has arrived at a critical point, which is written to all subsequent elements of the series. Otherwise, h_i is chosen to satisfy $\left\langle \nabla \tilde{F}(u_i), h_i \right\rangle_{H^1(\Omega)} < 0$, like $h_i = -\nabla \tilde{F}(u_i)$. u_{i+1} is computed and the method proceeds to the next step.

Much research has focused on the convergence of a gradient descent method to critical points in both finite and infinite dimensional settings (Curry, 1944; Polyak, 1963; Armijo, 1966; Byrd and Tapia, 1975; Karátson, 1999; Karátson et al., 2000;

Qian, 1999; Karátson et al., 2000; Penot, 2002; Gallego et al., 2015; Lee et al., 2016). Essentially, I wish to show that there exists $a > 0$ such that the gradient descent scheme converges for the functional in Equation 5.4, or that $\lim_{i\to\infty} u_i = u^*$, a critical point. Such methods always work for functionals whose gradient possesses the property of *uniform monotonicity* (Ljubič and Maistrovskif, 1970; Powell, 1971; Vajnberg, 1973; Berger, 1977), where there exists a $0 < c \in \mathbb{R}$ for the gradient of a functional given by ∇f such that

$$\langle \nabla f(u) - \nabla f(w), u - w \rangle \geq c \left|\left| u - w \right|\right|^2, \tag{5.83}$$

using the appropriate inner product and norm for all u, w in a Hilbert space. Uniform monotonicity implies a compactness property which has been shown to guarantee convergence (Izmailov and Tret'yakov, 1999) called the *Palais–Smale condition* (Palais and Smale, 1964; Mawhin and Willem, 2010). A continuously Fréchet differentiable functional $J \in C^1(\mathcal{H}, \mathbb{R})$ defined on Hilbert space $\mathcal{H}$ satisfies the Palais-Smale condition if every sequence $\{w_i\}_{i\in\mathbb{N}}$ such that $\{J(w_i)\}_{i\in\mathbb{N}}$ is bounded and the magnitude of the Fréchet derivative goes to zero in $\mathcal{H}$, or $\lim_{i\to\infty} ||DJ(u_i)||_{H^1} = 0$, has a convergent subsequence in $\mathcal{H}$.

The Palais-Smale condition in turn implies a weaker condition for convergence established by Smyrlis and Zisis (2004) which will guide the proof of convergence presented here, which relies upon a concept the authors call *condition* $(\mathcal{C})$.

Definition 5.4 (Condition $\mathcal{C}$)**.** *Let f be a C^1-functional defined on the real Hilbert space $\mathcal{H}$. The functional f is said to satisfy condition $(\mathcal{C})$, if for any closed bounded*

subset S *of* $\mathcal{H}$ *with*

$$||\nabla f(w)||_{\mathcal{H}} \neq 0 \ \ \forall w \in S, \tag{5.84}$$

then

$$\inf_{w \in S} ||\nabla f(w)||_{\mathcal{H}} > 0. \tag{5.85}$$

The authors note that every C^1 functional defined on $\mathbb{R}^n$ for $n \in \mathbb{N}$ satisfies condition $(\mathcal{C})$, and that while the Palais-Smale condition implies condition $(\mathcal{C})$, the opposite is not true.

The primary result of Smyrlis and Zisis (2004) is presented as Theorem 5.1.

Theorem 5.1. *Let* f *be a* C^1*-functional on Hilbert space* $\mathcal{H}$ *satisfying condition* $(\mathcal{C})$ *and attaining its minimum on the ball* $B(u^*, r)$ *with* $0 < r \in \mathbb{R}$ *centered at* u^**. Assume that* $B(u^*, r)$ *contains no critical points of* f *except for* u^* *and that the gradient operator* ∇f *is locally Lipschitz continuous. I choose the control parameters* $\rho_i \in (0, \infty), i \in N$*, so that*

$$0 < A \leq \rho_i \leq B < \frac{2}{K_f}, \tag{5.86}$$

where K_f *is the Lipschitz constant of* $\nabla f|_{B(u^*,r)}$ *and* A, B *are positive constants. Then there exists an open neighborhood* V *of* u^* *such that for each* $u_0 \in V$ *, the sequence* $\{u_i\}_{i \in \mathbb{N}}$ *defined by the iterative law*

$$u_{i+1} = u_i - \rho_i \nabla f(u_i), \tag{5.87}$$

for $i \in \mathbb{N}$*, converges to* u^**.*

Theorem 5.1 will be used to show that the gradient descent iterative law converges to a critical point if one exists. To show that a critical point exists I will use a result from Dacorogna (2004), presented as Theorem 5.2.

Theorem 5.2. *Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with Lipschitz boundary. Let $f \in C^0\left(\bar{\Omega} \times \mathbb{R} \times \mathbb{R}^n\right), f = f(\mathbf{x}, u, \xi)$, satisfy*

1. *$\xi \to f(\mathbf{x}, u, \xi)$ is convex for every $(\mathbf{x}, u) \in \bar{\Omega} \times \mathbb{R}$;*
2. *there exist $p > q \geq 1$ and $\alpha_1 > 0, \alpha_2, \alpha_3 \in \mathbb{R}$ such that*

$$f(\mathbf{x}, u, \xi) \geq \alpha_1 |\xi|^p + \alpha_2 |u|^q + \alpha_3, \ \ \forall (\mathbf{x}, u, \xi) \in \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^n.$$

Let

$$(P) \qquad \inf \left\{ I(u) = \int_\Omega f(\mathbf{x}, u(\mathbf{x}), \nabla u(\mathbf{x})) d\Omega \ : \ u \in u_0 + W_0^{1,p}(\Omega) \right\} = m$$

where $u_0 \in W^{1,p}(\Omega)$ with $I(u_0) < \infty$. Then there exists $\bar{u} \in u_0 + W_0^{1,p}(\Omega)$, a minimizer of (P).

Furthermore if $(u, \xi) \to f(x, u, \xi)$ is strictly convex for every $x \in \bar{\Omega}$, then the minimizer is unique.

Unfortunately, the functional in Equation 5.4 fails to meet condition $(\mathcal{C})$ in the infinite dimensional case. This ends up being because I can not place a lower bound $0 < c \leq F_{pp}(\mathbf{x}, u, \nabla u) \ \forall u \in H^1(\Omega)$. If $|\nabla u|$ becomes large, this term goes to zero and I will be unable to guarantee the gradient descent scheme converges to a solution. Additionally, I cannot guarantee the existence of a solution to the

minimization of Equation 5.4, because although it satisfies condition 1 of Theorem 5.2, it fails condition 2 as I can not place a lower bound such that $F(\mathbf{x}, u, \nabla u) \geq \alpha_1 |\nabla v|^p$ for $p \geq 2$, $\alpha_1 > 0$. This is closely related to why the functional fails to meet condition $(\mathcal{C})$.

In order to overcome this obstacle I take inspiration from the approach of Chambolle and Lions (1997) and Dobson and Vogel (1997) and modify the functional in Equation 5.4 to one for which a solution is guaranteed to exist and gradient descent convergence is guaranteed to occur,

$$\tilde{G}[v] = \int_{\Omega} e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2} |\nabla v|^2 \right) d\Omega, \tag{5.88}$$

where ϵ is a small positive number. As Dobson and Vogel (1997) note, this modification requires the resulting minimizer to be a $H^1(\Omega)$ function, which prevents it from having properties like possessing discontinuities along edges in two dimensions. This is not an issue for the physical application I use this picking scheme for, as the sort of semblance scans it operates on are created using RMS seismic velocities, which represent the integral in one dimension of of a L^2 interval velocity function which is assumed to vary smoothly laterally. Thus a solution of this problem is not expected to possess these sort of discontinuities.

I will not attempt to show that the solution is unique, because for $G(\mathbf{x}, v, p)$ to be convex in v, severe requirements are placed on $\alpha[v, \mathbf{x}]$, including

$$0 < \left(\alpha_v^2[v, \mathbf{x}] - \alpha_{vv}[v, \mathbf{x}] \right) e^{-\alpha[v,\mathbf{x}]} \; \forall \, v \in V, \, \mathbf{x} \in \Omega, \tag{5.89}$$

or equivalently,

$$\alpha_{vv}[v, \mathbf{x}] < \alpha_v^2[v, \mathbf{x}] \; \forall \, v \in V, \, \mathbf{x} \in \Omega. \tag{5.90}$$

Data commonly used for $\alpha[v, \mathbf{x}]$ have maxima in v where $\alpha_v[v, \mathbf{x}] = 0$. Thus, to fulfill strict convexity I would be requiring $\alpha_{vv}[v, \mathbf{x}] < 0 \; \forall \mathbf{x} \in \Omega$, $v \in V$, or for $\alpha[v, \mathbf{x}]$ to be strictly convex in v. This behavior is rare in the semblance scan data frequently used for $\alpha[v, \mathbf{x}]$, which are often non-convex in v, and also is in direct conflict with my assumption that $\bar{\alpha} \geq \alpha[v, \mathbf{x}] \geq 0$, since if α is bounded above it implies that $\alpha[v, \mathbf{x}] \to -\infty$ as $v \to \infty$.

I now define the integrand of Equation 5.88 to be

$$G(\mathbf{x}, v, p) = e^{-\alpha[v(\mathbf{x}), \mathbf{x}]} \left(\sqrt{\lambda^2 + |p|^2} + \frac{\epsilon}{2} |p|^2 \right), \tag{5.91}$$

and compute its partial derivatives

$$G_v(\mathbf{x}, v, p) = -\alpha_v[v(\mathbf{x}), \mathbf{x}] e^{-\alpha[v(\mathbf{x}), \mathbf{x}]} \left(\sqrt{\lambda^2 + |p|^2} + \frac{\epsilon}{2} |p|^2 \right), \tag{5.92}$$

and

$$G_p(\mathbf{x}, v, p) = e^{-\alpha[v(\mathbf{x}), \mathbf{x}]} \left(\frac{p}{\sqrt{\lambda^2 + |p|^2}} + \epsilon p \right). \tag{5.93}$$

These partial derivatives may be bounded by

$$|G_v(\mathbf{x}, v, p)| \leq L_{\alpha 1} \left(\lambda + |p| + \frac{\epsilon}{2} |p|^2 \right), \tag{5.94}$$

and

$$|G_p| \leq \sqrt{2} + \epsilon \, |p| \, . \tag{5.95}$$

I compute the second order partial derivatives

$$G_{vv}(\mathbf{x},v,p) = \left(\alpha_v^2[v(\mathbf{x}),\mathbf{x}] - \alpha_{vv}[v(\mathbf{x}),\mathbf{x}]\right) e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\sqrt{\lambda^2 + |p|^2} + \frac{\epsilon}{2}|p|^2\right), \quad (5.96)$$

$$G_{vp} = -\alpha_v[v(\mathbf{x}),\mathbf{x}] e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\frac{p}{\sqrt{\lambda^2 + |p|^2}} + \epsilon p\right), \quad (5.97)$$

and

$$G_{pp}(\mathbf{x},v,p) = e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\frac{\lambda^2}{\left(\lambda^2 + |p|^2\right)^{\frac{3}{2}}} + \epsilon\right). \quad (5.98)$$

The second order partial derivatives may be bounded derivatives using

$$|G_{vv}(\mathbf{x},v,p)| \leq \left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\left(\lambda + |p| + \frac{\epsilon}{2}|p|^2\right), \quad (5.99)$$

$$|G_{vp}(\mathbf{x},v,p)| \leq L_{\alpha 1}\left(\sqrt{2} + \epsilon|p|\right), \quad (5.100)$$

and

$$0 < \epsilon e^{-\tilde{\alpha}} \leq G_{pp} \leq \frac{1}{\lambda} + \epsilon. \quad (5.101)$$

Note that unlike the case for F_{pp} in Equation 5.18, I am able to place a lower bound on G_{pp} in Equation 5.101 that is greater than zero.

Now demonstrate that the functional given by Equation 5.88 has the properties required by Theorems 5.1 and 5.2 for a gradient descent scheme to converge for a minima if it does exist, and for a minima to exist.

Proposition 5.3. *The functional in Equation 5.91 is Fréchet differentialble. Its Fréchet derivative is*

$$D\tilde{G}(u)(\psi) = \langle \psi, G_v(\mathbf{x},u,\nabla u)\rangle_{L^2(\Omega)} + \langle \nabla\psi, G_p(\mathbf{x},u,\nabla u)\rangle_{L^2(\Omega)}. \quad (5.102)$$

Proof. The proof is identical to that of Proposition 5.1, except in place of Equation 5.33, the Gateaux derivative of $\tilde{G}$ is

$$\begin{aligned} d\tilde{G}(v;\psi) = &- \int_\Omega \psi \alpha_v[v(\mathbf{x}),\mathbf{x}] e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2} ||\nabla v||^2 \right) d\Omega \\ &+ \int_\Omega \nabla\psi \cdot \nabla v \, e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\frac{1}{\sqrt{\lambda^2 + |\nabla v|^2}} + \epsilon \right) d\Omega. \end{aligned} \tag{5.103}$$

Equation 5.103 is also defined for all $\psi \in H^1(\Omega)$ and $v \in V$, so Equation 5.88 is also Gateaux differentiable. Because $G(\mathbf{x}, v, p)$ is continuous in v and p just like $F(\mathbf{x}, v, p)$ the rest of the proof precedes identically, and I may state the Fréchet derivative of Equation 5.88,

$$D\tilde{G}(u)(\psi) = \langle \psi, G_v(\mathbf{x}, u, \nabla u) \rangle_{L^2(\Omega)} + \langle \nabla\psi, G_p(\mathbf{x}, u, \nabla u) \rangle_{L^2(\Omega)}. \tag{5.104}$$

□

Note that I may also define $\nabla\tilde{G}$ based on Equation 5.104 using the same argument as that employed for Equation 5.77

$$\nabla\tilde{G} = G_v(\mathbf{x}, u, \nabla u) - \nabla \cdot G_p(\mathbf{x}, u, \nabla u), \tag{5.105}$$

or equivalently

$$\begin{aligned} \nabla\tilde{G}(v) = &- \alpha_v[v(\mathbf{x}),\mathbf{x}] e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2} |\nabla v|^2 \right) \\ &- \nabla \cdot \left(e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left[\frac{1}{\sqrt{\lambda^2 + |\nabla v|^2}} + \epsilon \right] \nabla v \right), \end{aligned} \tag{5.106}$$

Proposition 5.4. *The Fréchet derivative of Equation 5.88 given by Equation 5.104 is locally Lipschitz continuous.*

Proof. This proof proceeds identically to that of Proposition 5.2 until the bounds of second order partial derivatives are applied in Equation 5.62. For the case of $D\tilde{G}$, for some $q \in B_{H^1(\Omega)}(0,R)$ this becomes

$$\begin{aligned} &||G_v(\mathbf{x},u,\nabla u) - G_v(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \\ &\qquad \leq \left|\left|\left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\left(\lambda + |\nabla q| + \frac{\epsilon}{2}|\nabla q|^2\right)|u-w|\right|\right|_{L^2(\Omega)} \\ &\qquad + \left|\left|L_{\alpha 1}\left(\sqrt{2} + \epsilon|\nabla q|\right)|\nabla u - \nabla w|\right|\right|_{L^2(\Omega)}. \end{aligned} \tag{5.107}$$

Using Cauchy-Schwartz and the absolute homogeneity of norms provides

$$\begin{aligned} &||G_v(\mathbf{x},u,\nabla u) - G_v(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \\ &\qquad \left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\left(\lambda + ||\nabla q||_{L^2(\Omega)} + \frac{\epsilon}{2}||\nabla q||_{L^2(\Omega)}^2\right)||u-w||_{L^2(\Omega)} \\ &\qquad + L_{\alpha 1}\left(\sqrt{2} + \epsilon||\nabla q||_{L^2(\Omega)}\right)||\nabla u - \nabla w||_{L^2(\Omega)}. \end{aligned} \tag{5.108}$$

Because $||\nabla q||_{L^2(\Omega)} < R$ I can write

$$\begin{aligned} &||G_v(\mathbf{x},u,\nabla u) - G_v(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \\ &\qquad \left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\left(\lambda + R + \frac{\epsilon}{2}R^2\right)||u-w||_{L^2(\Omega)} \\ &\qquad + L_{\alpha 1}\left(\sqrt{2} + \epsilon R\right)||\nabla u - \nabla w||_{L^2(\Omega)}. \end{aligned} \tag{5.109}$$

Using Equation 5.12 and the definition of the $H^1(\Omega)$ norm Equation 5.109 may again be bounded as

$$\begin{aligned} &||G_v(\mathbf{x},u,\nabla u) - G_v(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \leq \\ &\qquad \sqrt{2}\left[\left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\left(\lambda + R + \frac{\epsilon}{2}R^2\right) + L_{\alpha 1}\left(\sqrt{2} + \epsilon R\right)\right]||u-w||_{H^1(\Omega)}. \end{aligned} \tag{5.110}$$

I now apply a similar argument to G_p. This is quite similar to the arguments applied to F_p in the proof of Proposition 5.2 until the bounds of F_p are substituted

in Equation 5.68. In the case of G_p there exists a $q \in B_{H^1(\Omega)}(0,R)$ such that

$$\begin{aligned} ||G_p(\mathbf{x},u,\nabla u) - G_p(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \leq & \left|\left| L_{\alpha 1}\left(\sqrt{2}+\epsilon\left|\nabla q\right|\right)|u-w| \right|\right|_{L^2(\Omega)} \\ & + \left(\frac{1}{\lambda}+\epsilon\right)||\nabla u - \nabla w||_{L^2(\Omega)}, \end{aligned} \tag{5.111}$$

which may again be bound with

$$\begin{aligned} ||G_p(\mathbf{x},u,\nabla u) - G_p(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \leq & L_{\alpha 1}\left(\sqrt{2}+\epsilon\,||\nabla q||_{L^2(\Omega)}\right)||u-w||_{L^2(\Omega)} \\ & + \left(\frac{1}{\lambda}+\epsilon\right)||\nabla u - \nabla w||_{L^2(\Omega)}. \end{aligned} \tag{5.112}$$

Because $q \in B_{H^1(\Omega)}(0,R)$ $||\nabla q||_{L^2(\Omega)} \leq R$, and

$$\begin{aligned} ||G_p(\mathbf{x},u,\nabla u) - G_p(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \leq & L_{\alpha 1}\left(\sqrt{2}+\epsilon R\right)||u-w||_{L^2(\Omega)} \\ & + \left(\frac{1}{\lambda}+\epsilon\right)||\nabla u - \nabla w||_{L^2(\Omega)}. \end{aligned} \tag{5.113}$$

Using Equation 5.12 and the definition of the $H^1(\Omega)$ norm Equation 5.113 may be bounded as

$$\begin{aligned} &||G_p(\mathbf{x},u,\nabla u) - G_p(\mathbf{x},w,\nabla w)||_{L^2(\Omega)} \leq \\ &\qquad \sqrt{2}\left[L_{\alpha 1}\left(\sqrt{2}+\epsilon R\right)+\frac{1}{\lambda}+\epsilon\right]||u-w||_{H^1(\Omega)}. \end{aligned} \tag{5.114}$$

Introducing a bounding constant $L_{D\tilde{G}}$

$$L_{D\tilde{G}} = 2\left[\left(L_{\alpha 1}^2 + L_{\alpha 2}\right)\left(\lambda + R + \frac{\epsilon}{2}R^2\right) + L_{\alpha 1}\left(\sqrt{2}+\epsilon R\right)+\frac{1}{\lambda}+\epsilon\right], \tag{5.115}$$

using a similar argument to that employed for Equation 5.72 allows me to write

$$\left|D\tilde{G}(u)(\psi) - D\tilde{G}(w)(\psi)\right| \leq \frac{L_{D\tilde{G}}}{\sqrt{2}}||u-w||_{H^1(\Omega)}\left(||\psi||_{L^2(\Omega)} + ||\nabla\psi||_{L^2(\Omega)}\right). \tag{5.116}$$

Thus,

$$\left|D\tilde{G}(u)(\psi) - D\tilde{G}(w)(\psi)\right| \leq L_{D\tilde{G}} \left|\left|u - w\right|\right|_{H^1(\Omega)} \left|\left|\psi\right|\right|_{H^1(\Omega)}, \tag{5.117}$$

and therefore $D\tilde{G}$ is locally Lipschitz continuous with Lipschitz constant $L_{D\tilde{G}}$ given in Equation 5.115. As a corollary, $\tilde{G} \in C^1(H^1(\Omega), \mathbb{R})$ because any locally Lipschitz continuous function is also continuous. □

Proposition 5.5. *The functional $\tilde{G}$ defined by Equation 5.88 fulfills condition $(\mathcal{C})$.*

Proof. As was shown in the proof of Proposition 5.4, $D\tilde{G}$ is locally Lipschitz continuous, and therefore $\tilde{G} \in C^1(H^1(\Omega), \mathbb{R})$. In order to show the functional meets condition $(\mathcal{C})$ I therefore need to show that for any closed bounded subset S of $H^1(\Omega)$ with $||\nabla f(v)|| \neq 0 \ \forall v \in S$, that $\inf_{v \in S} ||\nabla f(v)|| > 0$. The proof presented here closely follows that employed in an example from Smyrlis and Zisis (2004), which is done by contradiction.

Let $S \subset H^1(\Omega)$ be closed and bounded such that $\left|\left|D\tilde{G}(v)\right|\right| \neq 0 \ \forall \, v \in S$, and suppose that condition $(\mathcal{C})$ does not hold for $\tilde{G}$. Then I may find a sequence $\{v_i\}_{n \in \mathbb{N}} \subset S$ such that $\left|\left|D\tilde{G}(v_i)\right|\right| \to 0$ as $i \to \infty$. Note that $\{v_i\}_{i \in \mathbb{N}} \subset S$ is bounded in the reflexive Banach space $H^1(\Omega)$, which is compactly embedded into $C(\Omega)$. Therefore, I may assume, passing to sub-sequences if necessary, that $\exists \ v \in H^1(\Omega)$ such that $v_i \xrightarrow{w} v$ in $H^1(\Omega)$ and $v_i \to v$ uniformly on Ω. Let

$$R_{\nabla v} = \sup_{i \in \mathbb{N}} \left|\left|\nabla v_i\right|\right|_{L^2(\Omega)} \tag{5.118}$$

and

$$R_v = \sup_{i \in \mathbb{N}} \left|\left|v_i\right|\right|_{C(\Omega)}, \tag{5.119}$$

both of which are finite because S is bounded. Then, allowing $\mu(\Omega) = \int_\Omega d\Omega$, which is also finite because Ω is bounded, I have

$$\sup_{i\in\mathbb{N}} ||G_v(\mathbf{x}, v_i, \nabla v_i)||_{L^2(\Omega)} \leq L_{\alpha 1}\left(\lambda\mu(\Omega) + R_{\nabla v} + \frac{\epsilon}{2}R_{\nabla v}^2\right) < \infty. \tag{5.120}$$

Additionally,

$$|G_p(\mathbf{x}, v_i, \nabla v)| \leq \sqrt{2} + \epsilon\,|\nabla v| \;\; \forall i \in \mathbb{N}. \tag{5.121}$$

Because $v_i \to v$ uniformly on Ω,

$$\int_\Omega G_v(\mathbf{x}, v_i, \nabla v_i)\,(v_i - v)\,d\Omega \to 0 \tag{5.122}$$

and using Lebesgue's dominated convergence theorem,

$$\lim_{i\to\infty} ||G_p(\mathbf{x}, v_i, \nabla v) - G_v(\mathbf{x}, v, \nabla v)||_{L^2(\Omega)} = 0. \tag{5.123}$$

Note that because $v_i \xrightarrow{w} v$ in $H^1(\Omega)$, $\nabla v_i \xrightarrow{w} \nabla v$ in $L^2(\Omega)$. Combining this with the result of Equation 5.123,

$$\lim_{i\to\infty} \int_\Omega G_p(\mathbf{x}, v_i, \nabla v)\cdot(\nabla v_i - \nabla v)\,d\Omega = 0. \tag{5.124}$$

Now, recalling that $D\tilde{G}(v)(\psi) = \left(D\tilde{G}(v), \psi\right)$ note that

$$\begin{aligned}\left(D\tilde{G}(v_i), v_i - v\right) = &\int_\Omega G_v(\mathbf{x}, v_i, \nabla v_i)\,(v_i - v)\,d\Omega \\ &+ \int_\Omega G_p(\mathbf{x}, v_i, \nabla v_i)\cdot(\nabla v_i - \nabla v)\,d\Omega,\end{aligned} \tag{5.125}$$

so

$$\begin{aligned}\int_\Omega [G_p(\mathbf{x}, v_i, \nabla v_i) - G_p(\mathbf{x}, v_i, \nabla v)]\cdot(\nabla v_i - \nabla v)\,d\Omega = \\ \left(D\tilde{G}(v_i), v_i - v\right) - \int_\Omega G_v(\mathbf{x}, v_i, \nabla v_i)\,(v_i - v)\,d\Omega \\ - \int_\Omega G_p(\mathbf{x}, v_i, \nabla v)\cdot(\nabla v_i - \nabla v)\,d\Omega.\end{aligned} \tag{5.126}$$

Therefore, using Equations 5.122 and 5.124,

$$\lim_{i\to\infty}\int_{\Omega}\left[G_p(\mathbf{x},v_i,\nabla v_i)-G_p(\mathbf{x},v_i,\nabla v)\right]\cdot(\nabla v_i-\nabla v)\,d\Omega=0. \tag{5.127}$$

Additionally, the lower bound established on $G_p p$ in Equation 5.101 combined with the mean value theorem entails that

$$\epsilon e^{-\bar{\alpha}}\left|p-q\right|\leq\left|(G_p(\mathbf{x},v_i,p)-G_p(\mathbf{x},v_i,q))\cdot(p-q)\right|. \tag{5.128}$$

Therefore

$$\lim_{i\to\infty}||\nabla v_i-\nabla v||_{L^2(\Omega)}=0, \tag{5.129}$$

and subsequently

$$\lim_{i\to\infty}||v_i-v||_{H^1(\Omega)}=0. \tag{5.130}$$

To complete the proof, recall that S is closed in $H^1(\Omega)$ and $D\tilde{G}$ is a continuous operator. Therefore $v\in S$ and $D\tilde{G}(v)=0$. But this is a contradiction with the assumption that $\left|\left|D\tilde{G}(v)\right|\right|\neq 0\ \forall\, v\in S$. Therefore, $\tilde{G}$ fulfills condition $(\mathcal{C})$. □

Proposition 5.6. *The function $G(\mathbf{x},u,p)$ given by Equation 5.91 is convex in p for every $(\mathbf{x},u)\in\bar{\Omega}\times\mathbb{R}$*

Proof. It is sufficient to show that $G_{pp}(\mathbf{x},u,p)\geq 0\ \forall p\in\mathbb{R}^n$, $(\mathbf{x},u)\in(\bar{\Omega}\times\mathbb{R})$. Using Equation 5.101, $G_{pp}(\mathbf{x},u,p)\geq\epsilon e^{-\bar{\alpha}}>0$.

□

Proposition 5.7. *There exist $m>n\geq 1$ and $\alpha_1>0,\alpha_2,\alpha_3\in\mathbb{R}$ such that*

$$G(x,u,p)\geq\alpha_1|p|^m+\alpha_2|u|^n+\alpha_3,\ \ \forall(x,u,p)\in\bar{\Omega}\times\mathbb{R}\times\mathbb{R}^n. \tag{5.131}$$

Proof. I begin by noting that $G(x,u,p) \geq \lambda e^{-\bar{\alpha}} > 0$, so I let $n = 0$, and $\alpha_2 = \alpha_3 = 0$. Therefore, setting $m = 2$, I wish to show there exists some α_1 such that

$$e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}\left(\sqrt{\lambda^2 + |p|^2} + \frac{\epsilon}{2}|p|^2\right) \geq \alpha_1 |p|^2 . \tag{5.132}$$

Because $\sqrt{\lambda^2 + |p|^2} \geq 0$ and $e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \geq e^{-\bar{\alpha}} > 0$, I can choose

$$\alpha_1 = \frac{\epsilon}{2} e^{-\bar{\alpha}}. \tag{5.133}$$

□

I am now able to present the primary result of this section, Theorem 5.3.

Theorem 5.3. *The functional $\tilde{G} : V \to \mathbb{R}^+$, defined as*

$$\tilde{G}(v) = \int_\Omega e^{-\alpha[v(\mathbf{x}),\mathbf{x}]}\left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2}|\nabla v|^2\right), \tag{5.134}$$

where

1. $0 < \lambda \in \mathbb{R}$,
2. $0 < \epsilon \in \mathbb{R}$,
3. $\Omega \subset \mathbb{R}^n$ *is a convex bounded set with Lipschitz boundary* $\partial\Omega = \Gamma$,
4. $V = \{v \in H^1(\Omega)$ such that $v \cdot \hat{\mathbf{n}} = 0$ for $\hat{\mathbf{n}} \in \Gamma\}$, *and*
5. $\alpha \in C^1(V \times \Omega, \mathbb{R})$ *such that* $0 \leq \alpha(v,\mathbf{x}) \leq \bar{\alpha}$ *is a Lipschitz continuous function in v such that $|\alpha_v| \leq L_{\alpha 1}$, and whose partial derivative in v is also Lipschitz continuous with $|\alpha_{vv}| \leq L_{\alpha 2}$,*

possesses a minimizer $v^* \in w_0 + H_0^1(\Omega)$*, where* $w_0 \in H^1(\Omega)$ *with* $\tilde{G}(w_0) < \infty$*. Assuming that there exists a ball* $B_{H^1(\Omega)}(v^*, r)$ *with* $r > 0$ *that contains no critical points of* $\tilde{G}$ *except for* v^**, then there exits an open neighborhood* U *of* v^* *such that for each* $v_0 \in U$ *the sequence* $\{v_i\}_{i\in\mathbb{N}}$ *defined by the iterative law*

$$v_{i+1} = v_i - \rho_i \nabla \tilde{G}(v_i), \tag{5.135}$$

where

$$\begin{aligned} \nabla \tilde{G}(v) = &- \alpha_v[v(\mathbf{x}), \mathbf{x}] e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left(\sqrt{\lambda^2 + |\nabla v|^2} + \frac{\epsilon}{2} |\nabla v|^2 \right) \\ &- \nabla \cdot \left(e^{-\alpha[v(\mathbf{x}),\mathbf{x}]} \left[\frac{1}{\sqrt{\lambda^2 + |\nabla v|^2}} + \epsilon \right] \nabla v \right), \end{aligned} \tag{5.136}$$

with control parameters ρ_i *selected such that*

$$0 < A \leq \rho_i \leq B < \frac{2}{L_{D\tilde{G}}}, \tag{5.137}$$

where A *and* B *are positive constants and* $L_{D\tilde{G}}$ *is the Lipschitz constant of* $\nabla \tilde{G}\Big|_{B_{H^1(\Omega)}(u^*,r)}$ *converges to* v^**.*

Proof. The proofs of Propositions 5.6 and 5.7 coupled with the Lipschitz continuity of $D\tilde{G}$ demonstrated in Proposition 5.4 demonstrates that Theorem 5.2 may be applied to $\tilde{G}$ and therefore that a minimizer exists. In the proof of Proposition 5.7, $m = 2$, and according to Theorem 5.2 a minimizer will lay in $W^{1,m}(\Omega) = H^1(\Omega)$. The proofs of Propositions 5.3 and 5.5 show that Theorem 5.1 may be applied to $\tilde{G}$, and that the outlined gradient descent scheme will converge to minimizer v^* assuming that there exists a ball $B_{H^1(\Omega)}(v^*, r)$ with $r > 0$ containing no critical points of $\tilde{G}$ except for v^*. □

Implementation

An iterative algorithm for minimizing Equation 5.88 to pick an optimal surface from a semblance-like pannel is implemented in the `Madagascar` software library (Fomel et al., 2013) using NumPy. Parallelism is achieved using Numba, and NumPy functions are rewritten as simple array-based object-free operations and performance tested using Numba on a TACC Stampede2 compute node with 48 Skylake cores for code optimization of each component helper function. The functional gradient $\nabla\tilde{G}(v)$, given by Equation 5.136 is computed using finite difference derivative operations and the Numba functions written for the task. The finite difference approach was chosen for numerical efficiency – representing the large semblance-like panels that are used for $\alpha[v(\mathbf{x}), \mathbf{x}]$ as FEniCS interpolated functions for finite element implementation was found to be overly expensive, and finite difference computations yielded satisfactory results at reduced computational cost. The calculated gradient is utilized with the ℓ-BFGS two loop recursion method for approximating the inverse Hessian and calculating a search direction outlined by Nocedal (1980). After experimentation on different data sets, I found that a memory size of the past 3 gradient computations for use in the approximation of the inverse Hessian achieved the most rapid cost convergence results. The step size is determined using that search direction with a golden-section search (Kiefer, 1953; Mordecai and Wilde, 1966). The method is allowed to iterate and update the model until either a maximum number of iterations are achieved, any step size in the search direction fails to decrease the cost more than a a set amount, or the L^2 magnitude of the calculated gradient drops blow a set threshold. Careful readers will notice a dimensional analysis ambiguity when calcu-

lating best fit surfaces from typical semblance panels – usually the first axis in these data are seismic traces, while other axes are spatial with units of distance. In order to convert derivatives in these different units into similar ones, spatial derivatives are multiplied by the velocity for the previous iteration so they have units of s^{-1}.

Convergence of this algorithm requires at least local convexity (Dennis and Moré, 1977; Nocedal, 1980; Li and Fukushima, 2001), which is be achieved by enforcing smoothness on $\alpha[v(\mathbf{x}), \mathbf{x}]$ using triangle smoothing filters (Claerbout, 1993). As a practical point, artificially large values in $\alpha[v(\mathbf{x}), \mathbf{x}]$ not related to high quality fit for the parameter being tested must be muted to prevent them from capturing the cost minimizing surface.

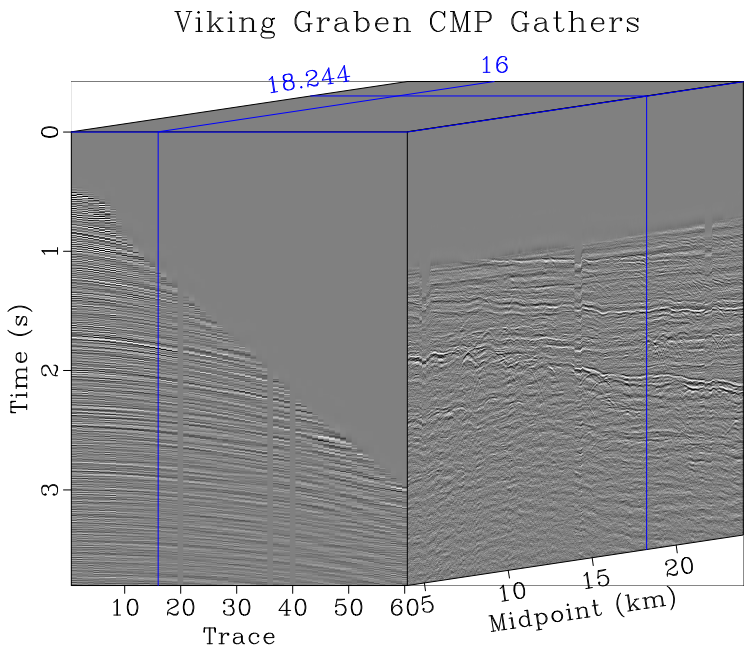
As noted in Theorem 5.3, convergence can only be guaranteed to a local minimizer given a starting model within a H^1 ball containing no other critical points of $\tilde{G}(v)$. In order for a unique minimizer to exist, the integrand of Equation 5.88 would need to be strictly convex in every v, ∇v combination for all $\mathbf{x}$ (Dacorogna, 2004), which places unrealistic burdens on α. In order to overcome the multimodality of this problem, I employ continuation, or graduated optimization. This involves smoothing and scaling the semblance-like volume a number of times. Smoothing makes the semblance volume more convex, and scaling increases the attraction of semblance highs for $v(\mathbf{x})$ surfaces. If sufficiently strong smoothing is applied the problem may become convex (Hazan et al., 2016). A constant gradient velocity field is used as the starting model on the most smoothed semblance volume, and the minimizing velocity surface is iteratively determined using the two loop recursion ℓ-BFGS scheme until convergence on that model is achieved. The minimizing surface from one level of smoothing

is then used as the starting model on the next less smoothed semblance volume. The method proceeds until finding the cost minimizing velocity on the least smoothed model. Because the method is allowed to operate on semblance-like volumes with different levels of smoothing, it is able to find the macro-trends for the best fit surface before searching out the finer components. This enables the method to behave more like a global optimizer, and enjoy applicability to problems beyond picking velocity surfaces from semblance-like panels, as experiments in subsequent sections illustrate.

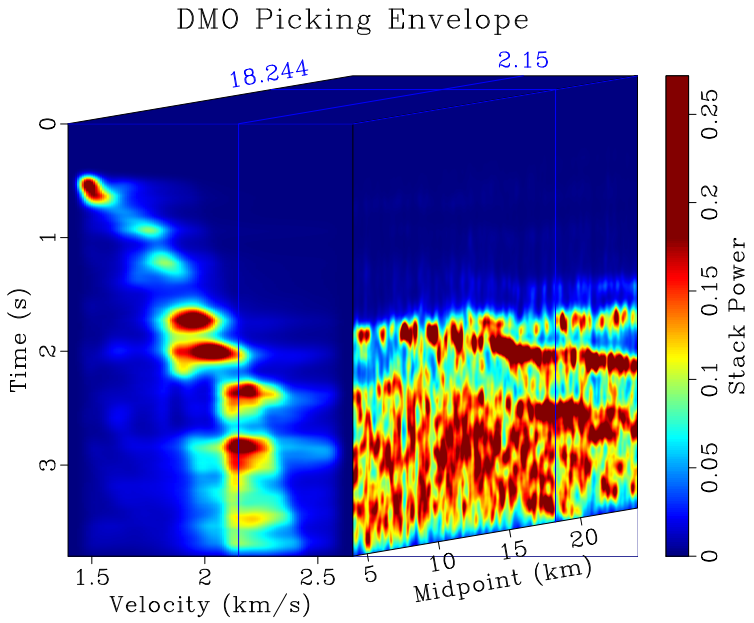
FIELD DATA EXAMPLES

Viking Graben

I apply the method to a field data example from the Viking Graben. Figure 5.1(a) contains a set of common midpoint (CMP) gathers which were preprocessed to have multiples removed using the parabolic Radon transform. I apply the DMO method from Fowler (1988) to to generate a series of constant velocity stacks, from which the stack power or envelope is computed. This will serve as $\alpha[v, \mathbf{x}]$. High values in the envelope not corresponding to physically plausible velocities are muted. Smoothing and scaling of increasing strength is applied to the muted volume for use in continuation. The least smoothed version is shown in Figure 5.1(b). I generate 10 volumes of increasingly aggressive smoothing and scaling and apply the ℓ-BFGS variational velocity picking algorithm with continuation using those volumes. The initial continuation level is shown in Figure 5.2(a). The initial model shown in solid black is a linear $v(t)$ spanning the range of scanned velocities and record times. The final model of this continuation level is shown in dashed white. The fifth continu-



(a)



(b)

Figure 5.1: (a) Common midpoint gathers from the Viking Graben; (b) Constant velocity DMO stack power.

ation level is shown in Figure 5.2(b), the ninth in Figure 5.2(c), and the tenth and final level in Figure 5.2(d). In each of these panels the solid black line represents the stating model for the continuation level, which was the final model of the previous one, and the dashed white line is the final model. Figure 5.2(d) also features the linear initial model in solid red, which is the same as the starting model in Figure 5.2(a), and the picking output without continuation in dashed green. This model is generated applying the ℓ-BFGS picking approach only on this level of smoothing. This non-continuation final model is shown in Figure 5.3(a), and the final model with continuation is displayed in Figure 5.3(b). The cost functional in Equation 5.88 is evaluated for each update of both methods using the least smoothed stack power volume in Figure 5.1(b) and plotted in Figure 5.4. The blue line plots the costs of the method without continuation and the red line plots the costs with continuation.

I use the two velocity fields to DMO stack and then Kirchhoff migrate the data. Figure 5.5(a) contains the non-continuation velocity model image, and the continuation velocity model image is shown in Figure 5.5(b). I also compute the Dix, or interval, velocity of the two models. This is a representation of local velocity of material at each region in the subsurface. Picked velocities were RMS velocities, corresponding to a mean velocity value over the complete ray path. The non-continuation Dix velocity is plotted in Figure 5.6(a), and the continuation Dix velocity is shown in Figure 5.6(b). These Dix velocities are used for conversion from the time to the depth domain following the manner of Sripanich and Fomel (2018). The depth stretched Dix velocity and image for the non-continuation velocity model are visible in Figures 5.7(a) and 5.7(b) respectively. The depth stretched Dix velocity

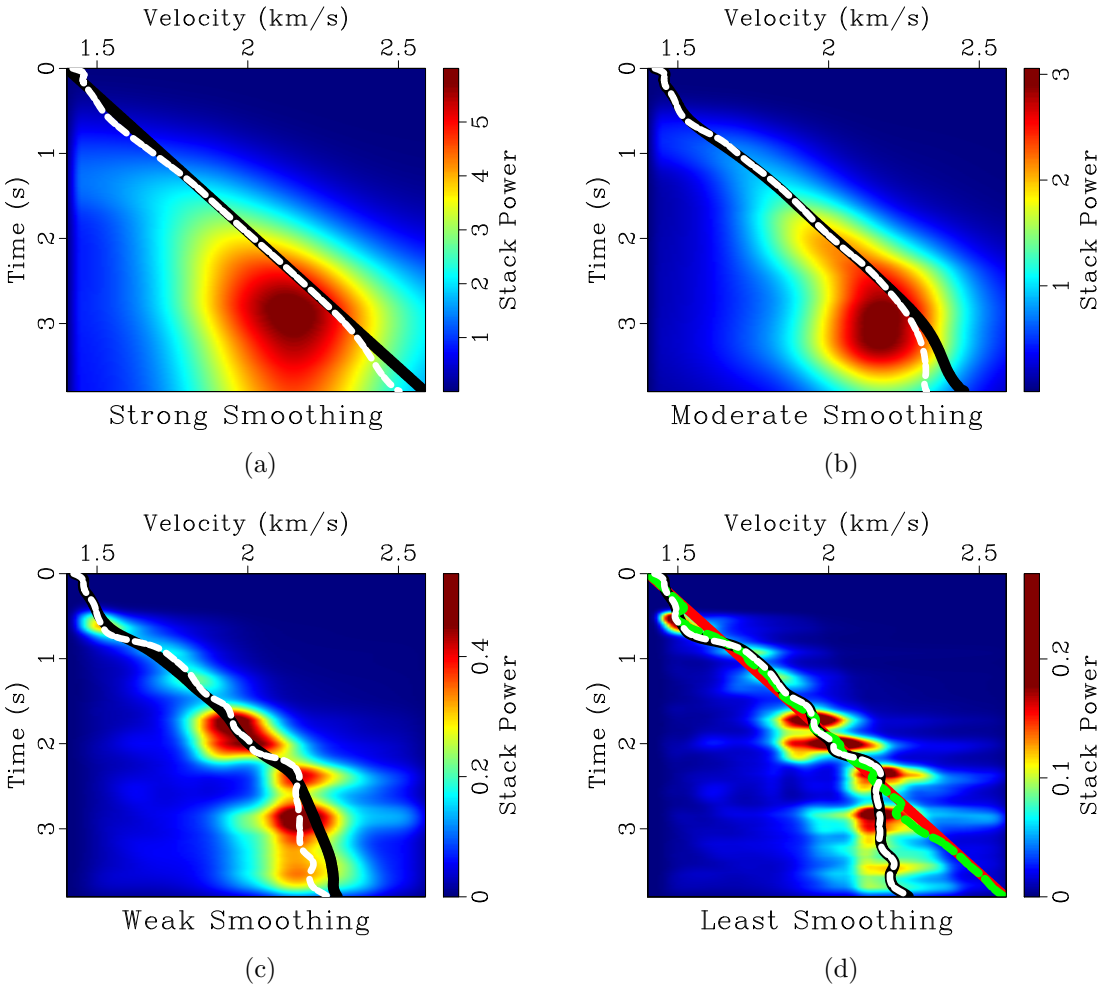
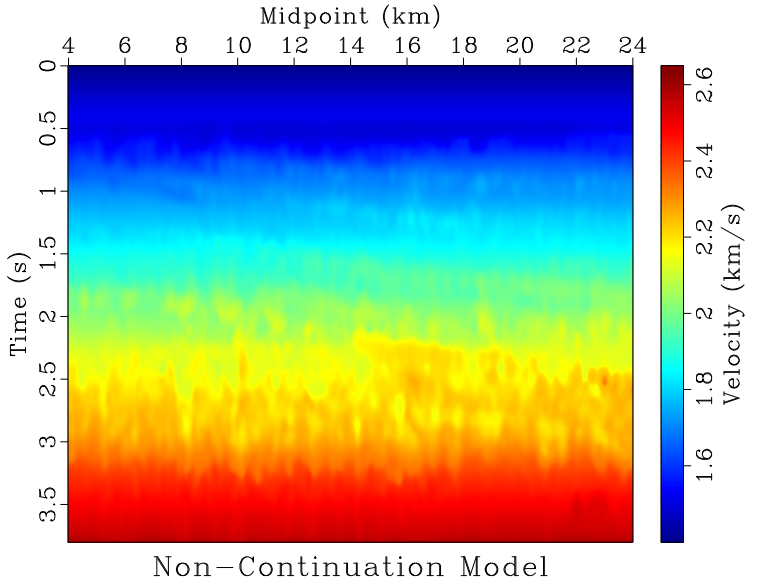
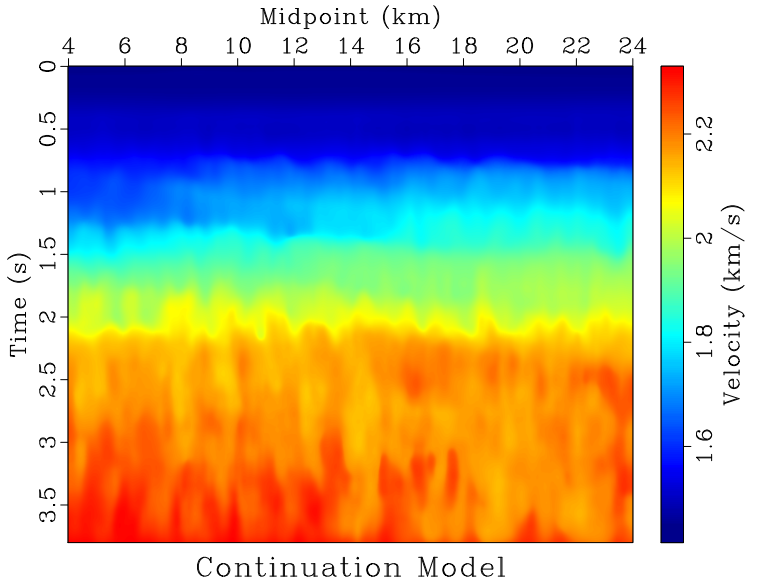


Figure 5.2: Illustration of continuation picking for midpoint at 18.244 km. Solid black line plots starting model for each level, dashed white line shows final model. (a) Initial continuation level with strongest smoothing; (b) Middle continuation level with moderate smoothing; (c) Second to last continuation level with weak smoothing; (d) Final continuation level also plotting the initial (solid red) and final models (dashed green) for picking without continuation.



(a)



(b)

Figure 5.3: (a) Velocity model determined by variational picking algorithm without continuation; (b) Velocity model determined by variational velocity picking algorithm utilizing continuation.

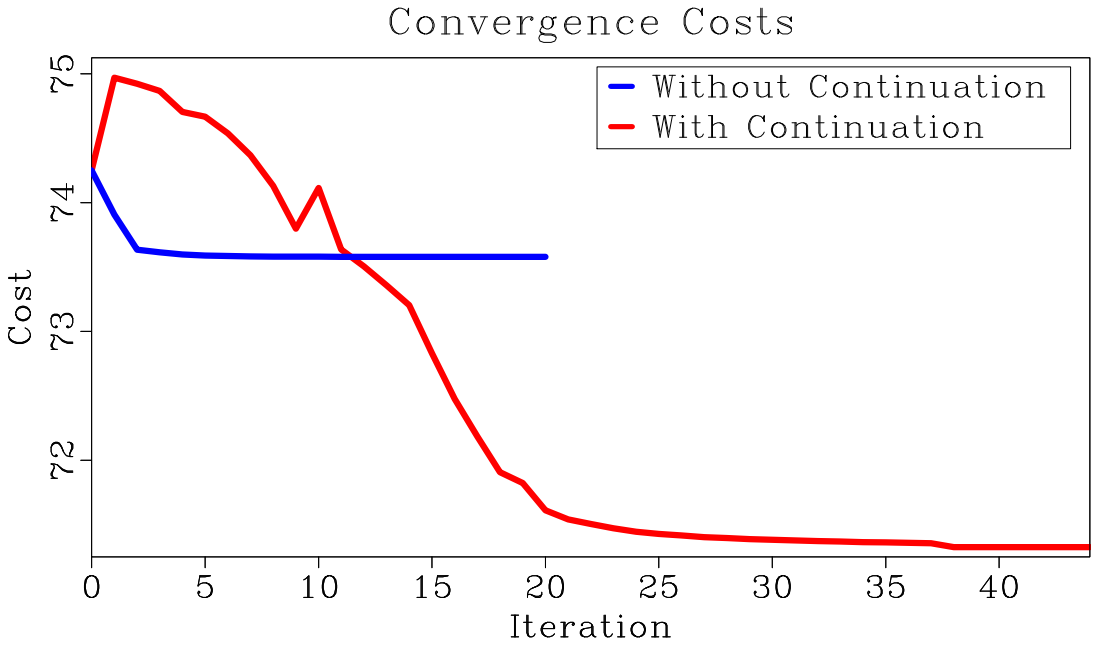
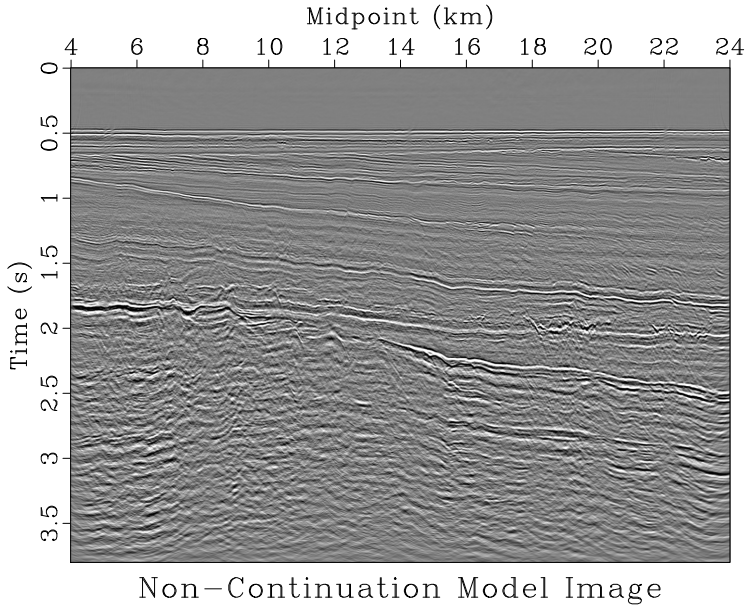


Figure 5.4: Velocity model cost, $\tilde{G}(v_i)$, computed using the least smoothed stack power volume visible in Figure 5.1(b).

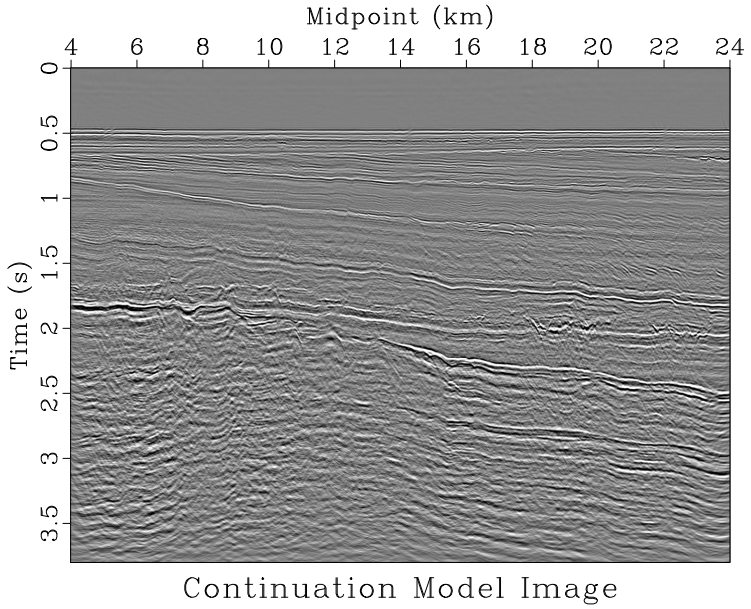
for the continuation model is plotted in Figure 5.8(a) and the corresponding depth stretched image in Figure 5.8(b).

Notice that continuation behaves as a more global optimization scheme, enabling the solution to avoid local minima. In fact, on the first continuation level shown in Figure 5.2(a) the cost actually increases when calculated on the least smoothed stack power volume.This enables the continuation velocity model to find the dominant trend for times greater than 2.5 s where it deviates from the starting model, and produce a final model with a low associated cost. Although the velocity field found without continuation is able to track the dominant trend and mesh with the continuation model fairly well for times less than 2.5 s, as shown in Figure 5.2(d). Instead, it encounters a local minimum with higher cost than the one found using continuation. This behavior can also be seen in Figures 5.3(a) and 5.3(b), which match reasonably well until 2.5 s, below which the model without continuation has higher velocities (note the differing scale bars).

Differences resulting from the two velocity models can be see in the migrated images, Figures 5.5(a) and 5.5(b), which are similar until 2.5 s. Below that time, reflectors become more poorly defined in the non-continuation model image, while in the continuation image coherent reflection events are visible, particularly between 15 and 24 km. This more coherent image is related to the continuation velocity following the dominant trend in the stack power volume. In the two Dix velocities, Figures 5.6(a) and 5.6(b) accentuate the differences between the models (note the difference in scale bars). This is because Dix, or interval, velocity is based on the derivative of RMS velocity that is used in DMO stacking and thus the generation of the

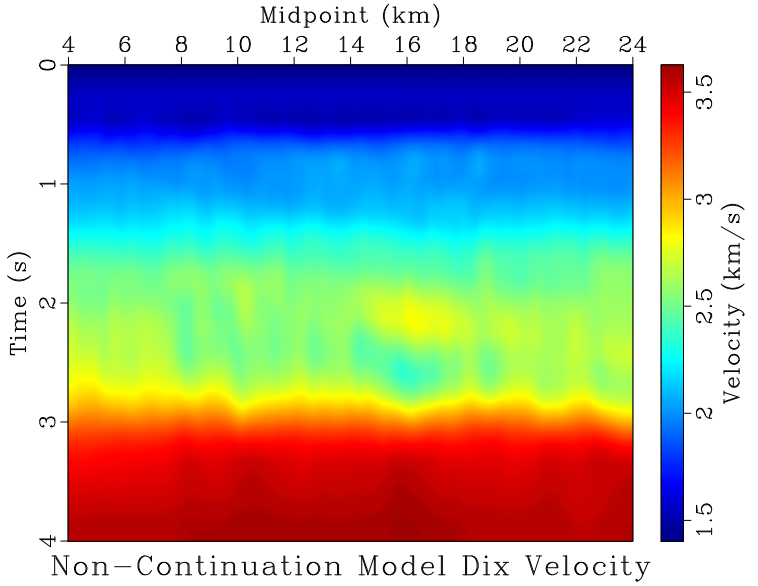


(a)

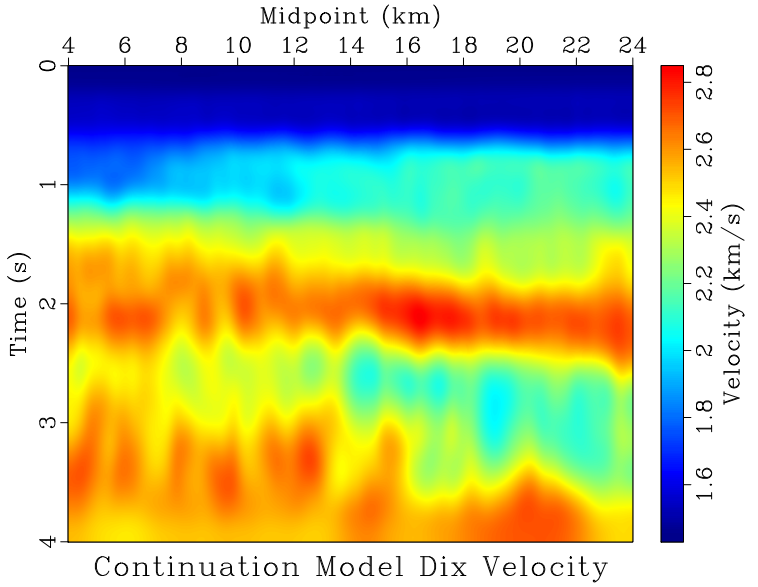


(b)

Figure 5.5: Kirchhoff time migrated images following DMO stacking using (a) the velocity model without continuation in Figure 5.3(a); (b) the continuation velocity model in Figure 5.3(b).

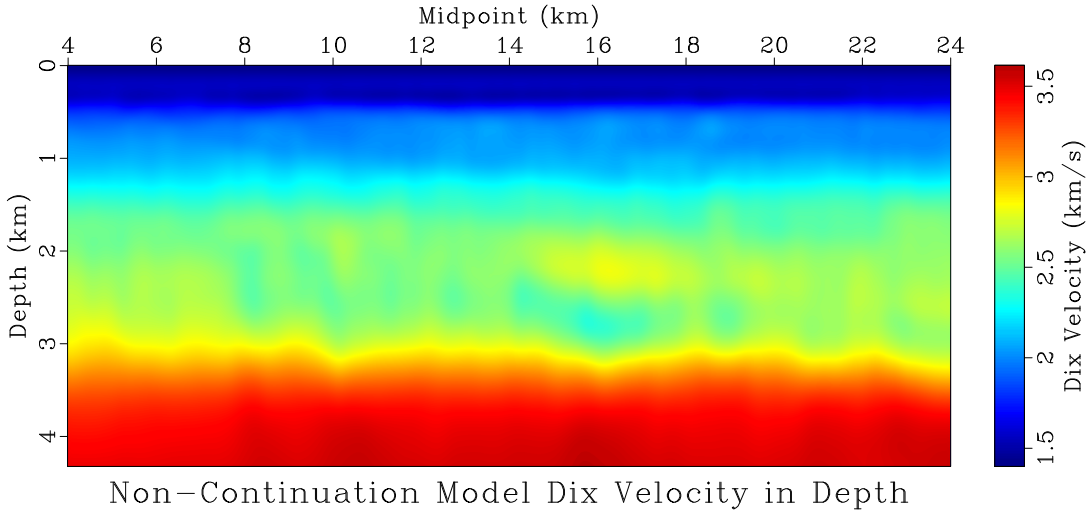


(a)

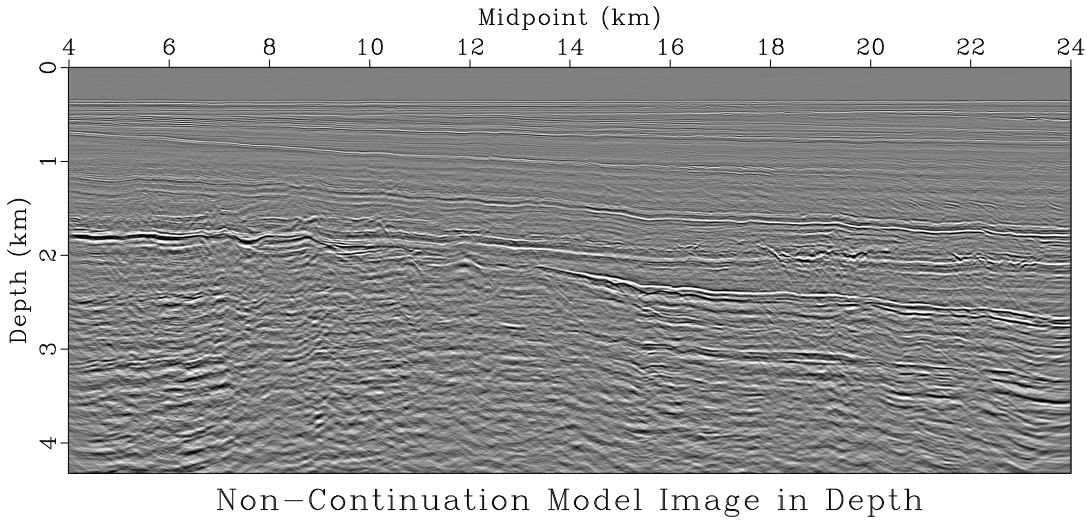


(b)

Figure 5.6: Dix velocities corresponding to (a) the velocity model without continuation in Figure 5.3(a); (b) the continuation velocity model in Figure 5.3(b).

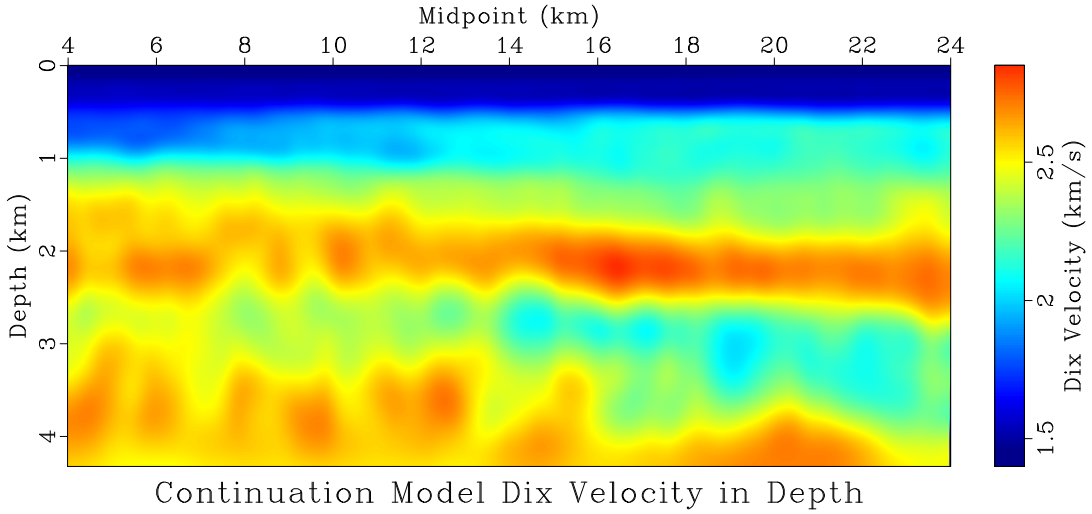


(a)



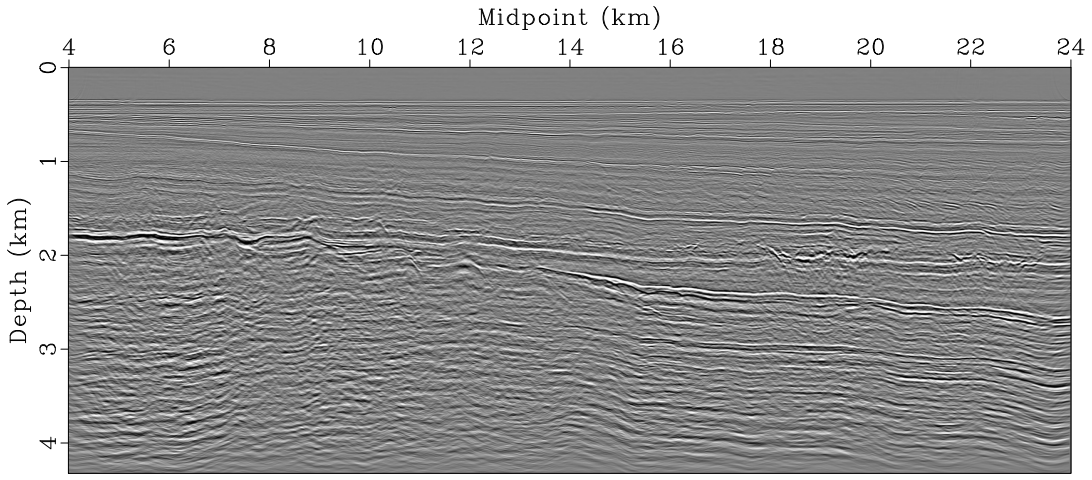
(b)

Figure 5.7: (a) Dix velocity corresponding to non-continuation model in Figure 5.3(a) transformed to the depth domain; (b) Non-continuation model image in Figure 5.5(a) transformed to the depth domain.



Continuation Model Dix Velocity in Depth

(a)



Continuation Model Image in Depth

(b)

Figure 5.8: (a) Dix velocity corresponding to continuation model in Figure 5.3(b) transformed to the depth domain; (b) Continuation model image in Figure 5.5(b) transformed to the depth domain.

α volume used in this experiment. Examining the Dix velocity in the depth domain for the non-continuation model Figure 5.7(a), and the depth domain Dix velocity for the continuation mode, Figure 5.8(a), this departure between the models corresponds to a well-defined velocity anomaly beginning near 2.25 km depth in the continuation Dix velocity. There is a slight anomaly present in the non-continuation Dix velocity model beginning around 2.5 km, but it is less well defined. Looking at the non-continuation seismic image transformed to the depth domain, Figure 5.7(b) and its continuation counterpart, Figure 5.8(b), notice that as was the case in the time domain images, the continuation image in the depth domain features reflectors that are more laterally coherent for depths greater than 2.5 km. Because the non-continuation Dix velocity model tends to have higher velocities than the continuation velocity model, reflectors in the non-continuation model image in depth tend to occur at greater depths than in the continuation velocity model. Because there is no corresponding high for those velocities in the DMO picking envelope, Figure 5.1(b), one may safely assume that the depths these reflectors have been mapped to are incorrect. Thus, using the continuation velocity model would lead to superior subsurface characterization, as seismic reflectors and the changes in lithology they correspond to are mapped to depths more representative of their position within the Earth.

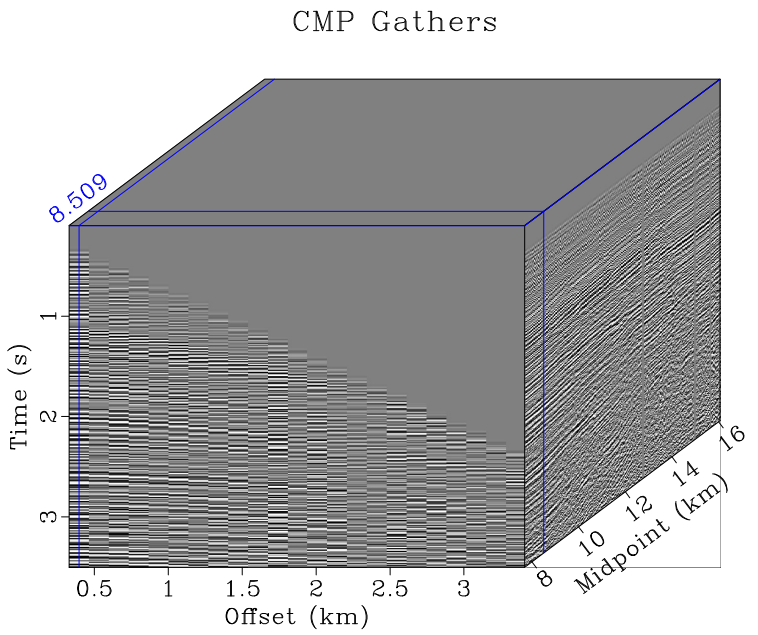
Gulf of Mexico

I now use a field dataset from the Gulf of Mexico to illustrate how the continuation approach coupled with the iterative algorithm for velocity picking proposed here can act like a global minimizer, significantly reducing the dependance of the

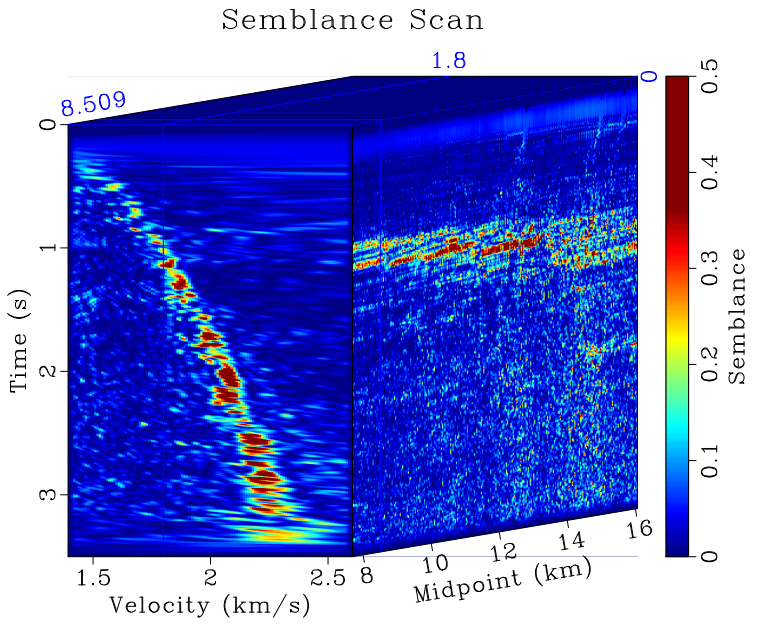
final model on the starting model. Figure 5.9(a) contains a set of common midpoint (CMP) gathers. A normal moveout (NMO) velocity scan is performed on the gathers to calculate semblance over a range of plausible NMO velocities and shown in Figure 5.9(b). Ten levels of increasingly aggressive smoothing and scaling are applied to this NMO scan to create an $\alpha[v(\mathbf{x}), \mathbf{x}]$ for each of ten continuation levels.

To illustrate how the continuation approach reduces the dependance of the final model on the starting model, I generate 125 constant gradient velocity models which are clipped to the range of the velocity scan. A random selection of nine of these are shown in Figures 5.10(a) through 5.10(i). These 125 starting models are used with the continuation picking approach over the ten smoothness levels, and their cost convergence as determined on the least smoothed semblance volume is plotted in Figure 5.11(a). If convergence is achieved before the maximum possible number of iterations, the final cost is repeated in these plots thru the maximum iteration number. The lowest const achieved is plotted as a dashed black line in that plot. For comparison I also use the collection as starting models for iterative picking on the least smooth $\alpha[v(\mathbf{x}), \mathbf{x}]$ without continuation. That cost convergence is plotted in Figure 5.11(b). Again, the lowest cost achieved by the continuation picking approach is plotted as a dashed black line and if convergence is achieved before the maximum possible number of iterations, the final cost is repeated in these plots thru the maximum iteration number.

The final continuation model with the lowest cost ($\tilde{G}(v) = 129.598$) is plotted in Figure 5.12(a), and the continuation model with the highest cost ($\tilde{G}(v) = 129.925$) is shown in Figure 5.12(b). The lowest cost ($\tilde{G}(v) = 130.278$) non-continuation final



(a)



(b)

Figure 5.9: (a) CMP gathers from a Gulf of Mexico field dataset; (b) NMO velocity scan for the CMP gathers in Figure 5.9(a).

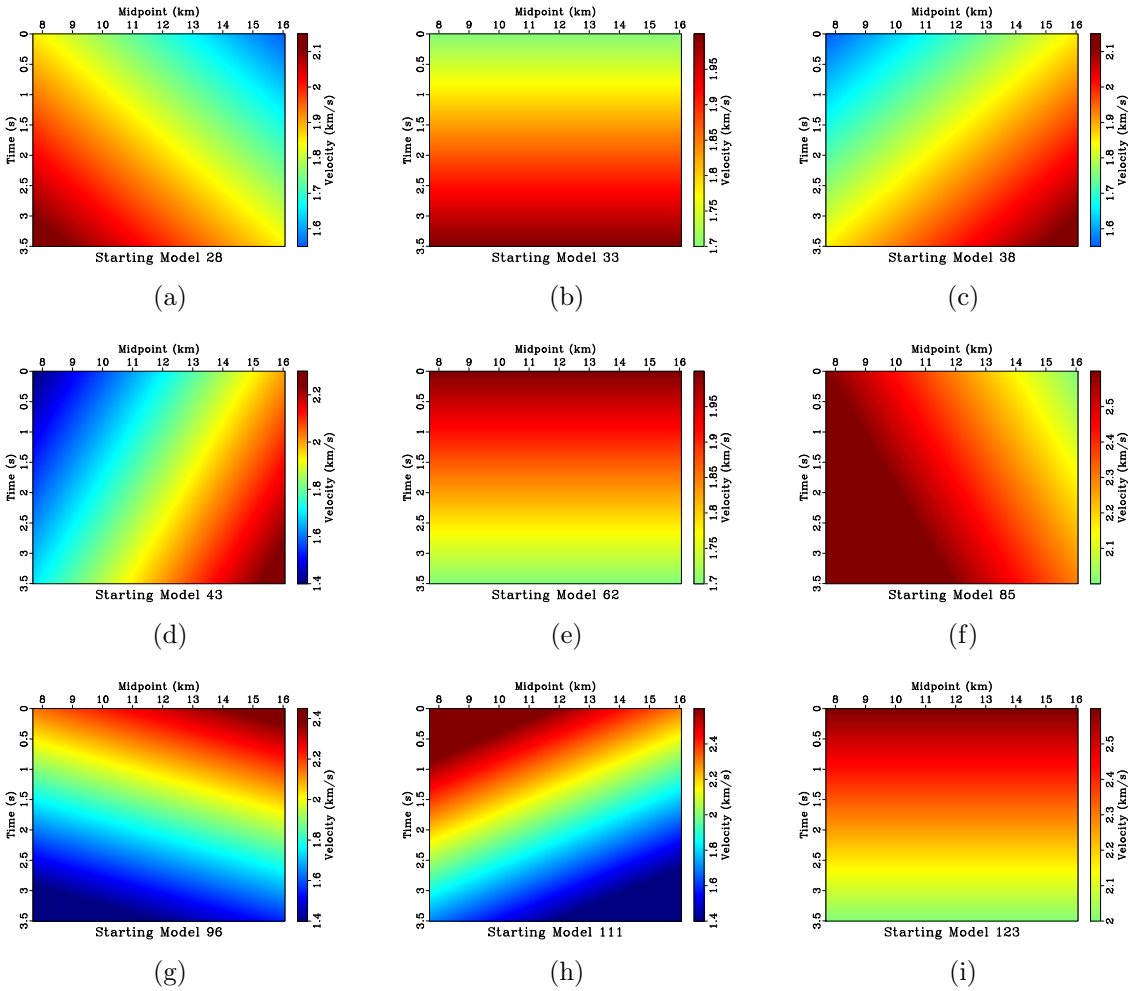
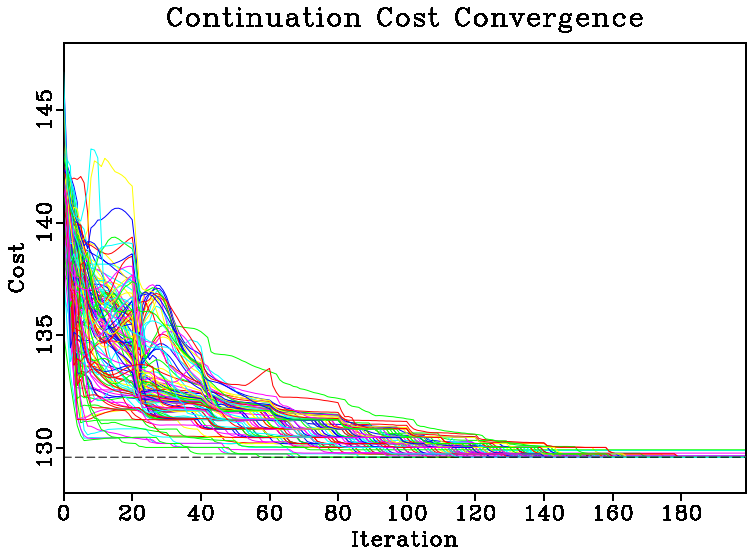
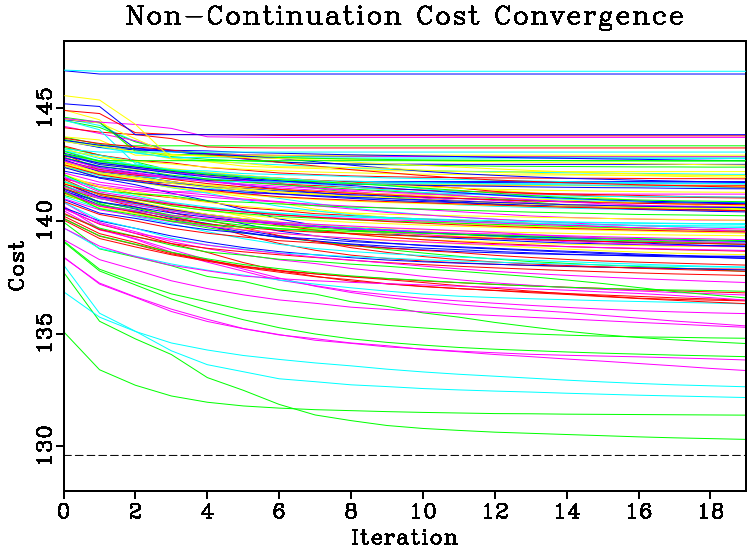


Figure 5.10: Random selection of nine constant gradient velocity models from the 125 used as starting models.



(a)



(b)

Figure 5.11: Cost convergence $\tilde{G}(v_i)$ for the 125 constant gradient velocity models calculated on the least-smoothed semblance volume (a) using the continuation approach; (b) without continuation. The lowest cost achieved using continuation is plotted as a dashed black line.

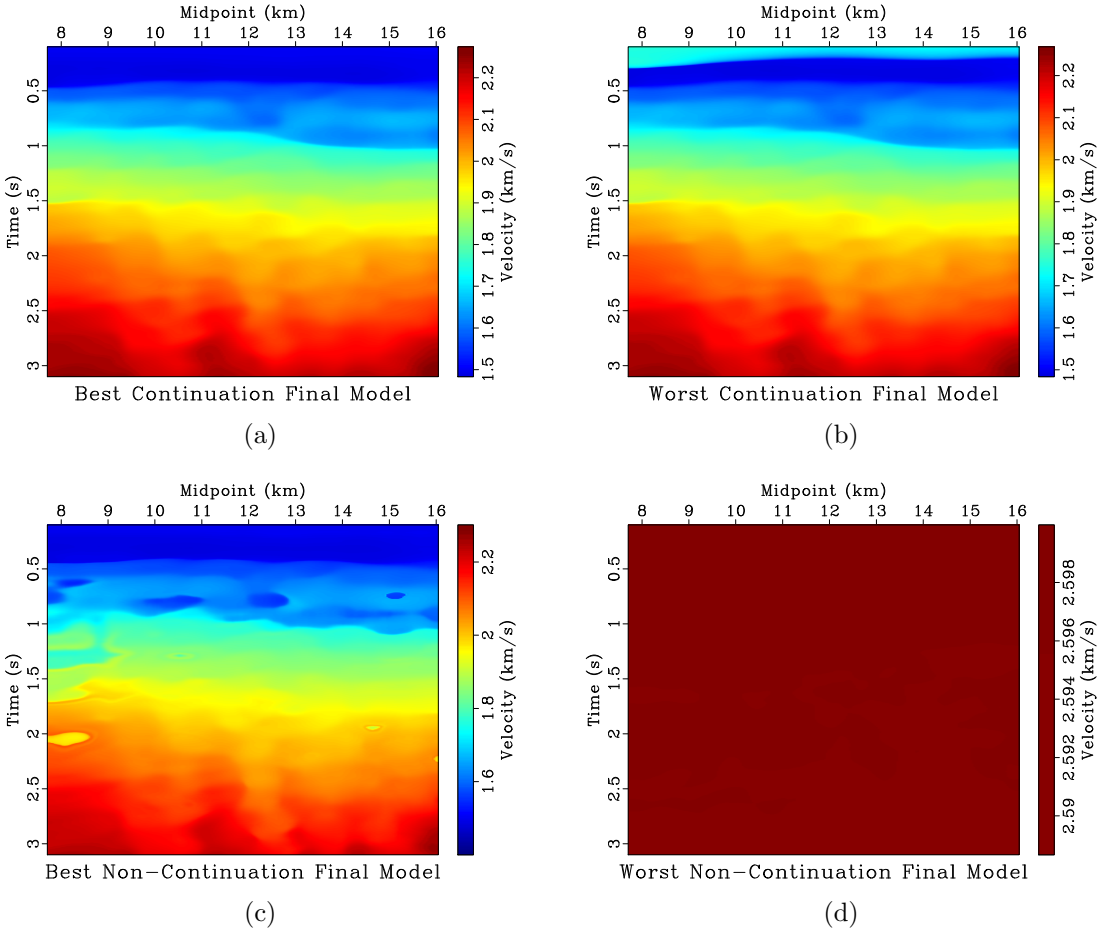
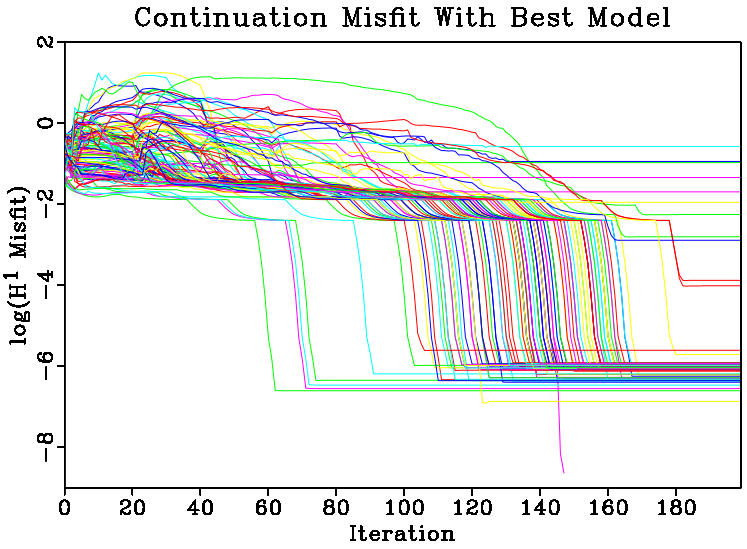


Figure 5.12: Lowest and highest cost final models from velocity picking using the 125 constant gradient starting models: (a) lowest cost continuation final model; (b) highest cost continuation final model; (c) lowest cost non-continuation final model; (d) highest cost non-continuation final model.

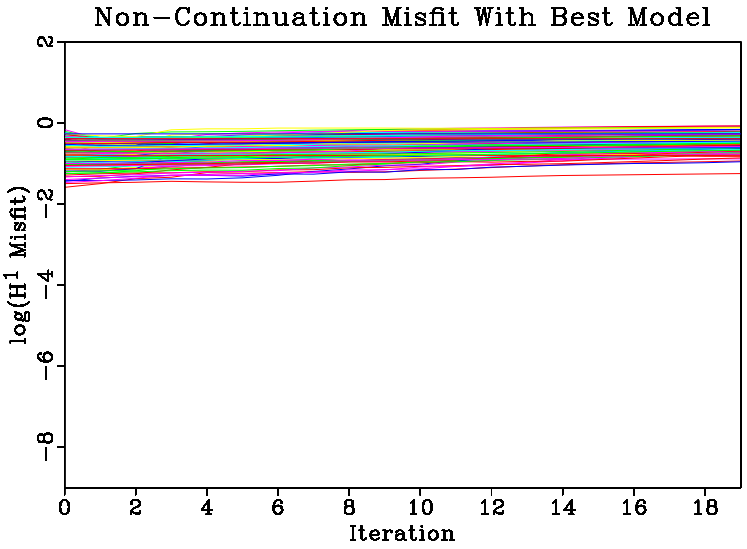
model can be seen in Figure 5.12(c), and the highest cost ($\tilde{G}(v) = 146.644$) non-continuation final model is displayed in Figure 5.12(d). To show how the continuation approach tends to produce similar final models, I compute the H^1 difference between the lowest cost continuation final model shown in Figure 5.12(a) and every update for every model and plot them in log scale, as shown for the continuation approach in Figure 5.13(a), and non-continuation approach in Figure 5.13(b). If convergence is achieved before the maximum possible number of iterations, the final H^1 misfit is repeated in these plots thru the maximum iteration number. Notice that in the continuation H^1 convergence plot shown in Figure 5.13(a), the lowest cost model can be in fuchsia descending to an isolated end point around -9 near iteration 145. This is because in the next iteration it achieves the lowest cost final model, and thus the H^1 misfit is zero, the log of which is undefined.

The lowest cost continuation velocity model, which is subsequently referred to as the "best model", is used for seismic processing of this dataset. The left panel of Figure 5.14 visualizes the best model overlaid on the semblance scan for a CMP gather at 8.5 km. The right panel illustrates that gather after NMO correction with the best model has been applied. Figure 5.15(a) contains stacked data following NMO correction using the best model. Diffraction data are extracted from that stack using plane-wave destruction filters and displayed in Figure 5.15(b). The best model is used for Kirchhoff time migration on the best model on NMO stack, whose image is shown in Figure 5.16(a), and the diffraction data as shown in Figure 5.16(b).

The continuation approach to picking velocities is able to overcome local minima and generate final models of similar cost and appearance relative to those created



(a)



(b)

Figure 5.13: log plots of H^1 difference between model updates for all 125 constant gradient starting models at each iteration and lowest cost continuation final model shown in Figure 5.12(a) for: (a) continuation picking; (b) non-continuation picking.

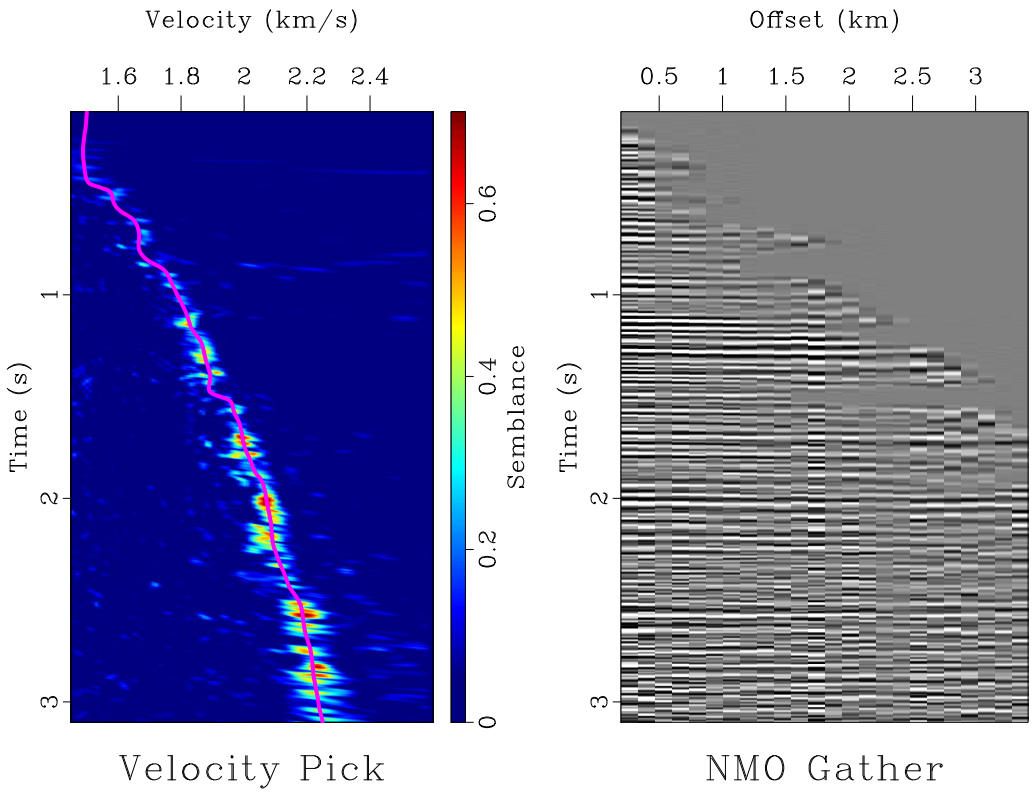
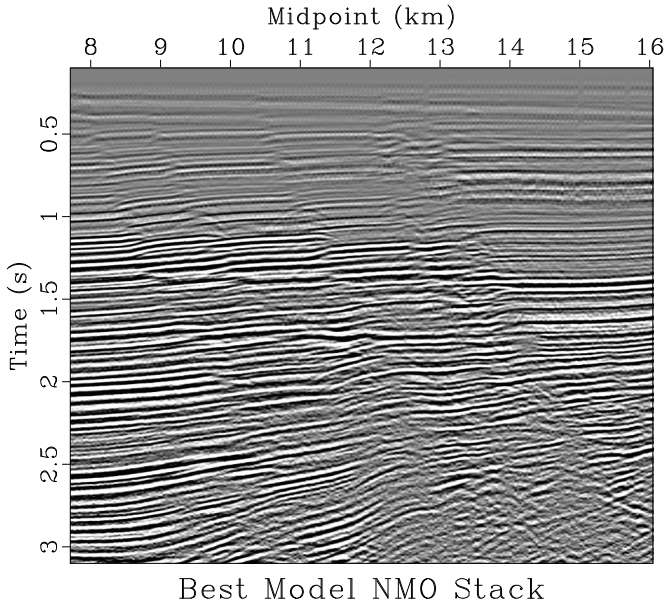
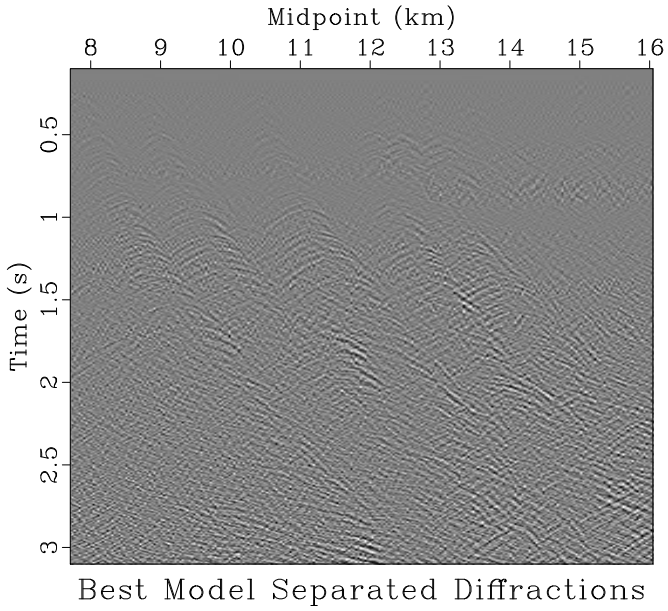


Figure 5.14: Illustration of the effects of the picking algorithm on a CMP centered at 8.5 km. Left panel contains a semblance scan for the CMP gather centered here overlaid by lowest cost final continuation velocity model from Figure 5.12(a) in solid fuchsia. Right panel contains the NMO corrected gather using that lowest cost final velocity model.

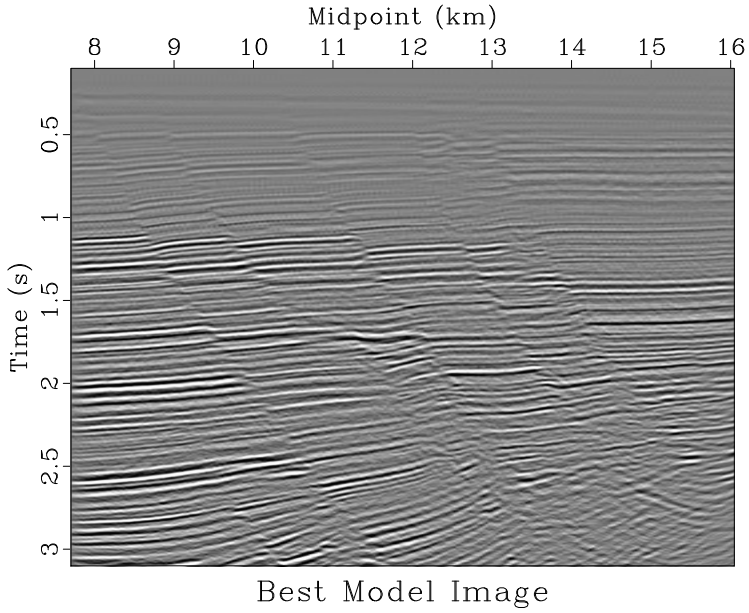


(a)

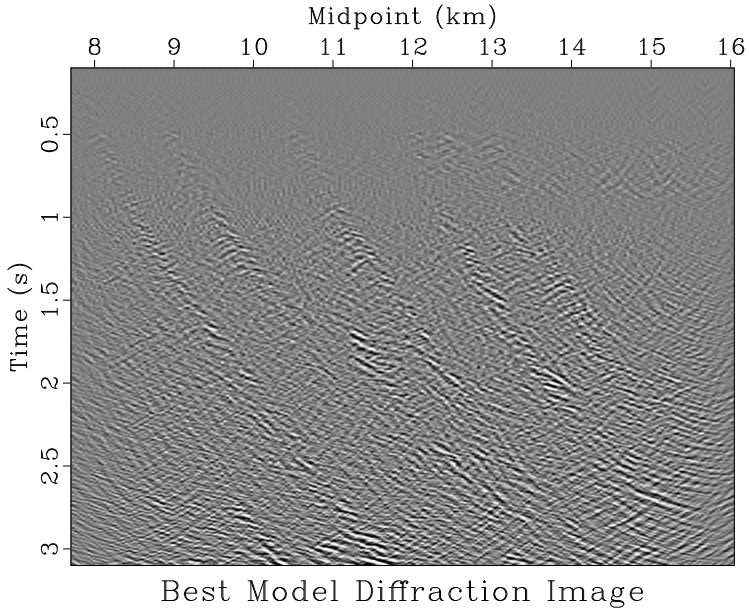


(b)

Figure 5.15: (a) NMO corrected stack generated using the lowest cost continuation model shown in Figure 5.12(a); (b) Diffraction data extracted from the NMO stack in Figure 5.15(a) using plane-wave destruction filters.



(a)



(b)

Figure 5.16: Kirchhoff time images generated using the lowest cost continuation model shown in Figure 5.12(a) corresponding to (a) the complete NMO stack data displayed in Figure 5.15(a); (b) the diffraction data shown in Figure 5.15(b).

without continuation. As can be seen in the cost convergence plots for the continuation approach, Figure 5.11(a), and the cost convergence plot without continuation, Figure 5.11(b), the continuation approach leads to uniformly low cost models relative to the approach without continuation. Indeed, the lowest cost model achieved with continuation possesses a cost of $\tilde{G}(v) = 129.598$ while the highest cost model possesses a cost of $\tilde{G}(v) = 129.925$. These costs are significantly lower than the starting model costs for the 125 constant gradient models, which range from roughly $\tilde{G}(v_o) = 135$ to $\tilde{G}(v_o) = 148$. Carefully examining Figure 5.11(b) there are only four final models of the 125 whose final cost is visibly different from the lowest cost model. Comparing this result to the cost convergence without continuation in Figure 5.11(b), the lowest cost non-continuation model has a higher cost ($\tilde{G}(v) = 130.278$) than the highest cost continuation model. There is also a significantly larger spread in final costs.

The continuation approach tends to generate to models that appear similar. The lowest cost continuation model, plotted in Figure 5.12(a), and the highest cost non-continuation model in Figure 5.12(b) are essentially the same below 0.3 s. This makes sense, because that is the approximate location of the seafloor. Examining the semblance scan used to create these models in Figure 5.9(b) there is no meaningful semblance information above this time to guide the picked velocity, since no reflections exist there to guide the NMO scan. The lowest cost non-continuation model in Figure 5.12(c) has a reasonably similar appearance to the lowest cost continuation final model, although it possesses some anomalous "blobs" that do not appear geological. The highest cost non-continuation final model in Figure 5.12(d) bears little, if any, resemblance to the lowest-cost continuation final model, and does not appear par-

ticularly geological or informative of the subsurface. Examining the H^1 convergence plot of the continuation approach, all but eleven of the 125 initial constant gradient velocity models achieve misfits with the lowest cost model lower than 10^{-5} (note that these calculations were performed using single precision arithmetic, so some of this misfit may be due to rounding error). The majority of those models follow paths that significantly decrease their H^1 error as the number of iterations increase, so they move "toward" the lowest-cost model. Examining those eleven models failing to fall within that range, the misfit is primarily borne in the areas less than 0.3 s where no reflections exist to guide the models. Conversely, the H^1 error without continuation has no obvious decreasing trend, rather their error actually tends to increase indicating they are moving "away" from the lowest cost model. No non-continuation models attain misfit less than 10^-2, and the lowest misfit is actually attained by a starting model.

The semblance scan overlaid by lowest cost continuation velocity in the left panel of Figure 5.14 shows a picked velocity that tracks the dominant trend in the semblance scan. The right panel of that figure shows the corresponding NMO corrected gather. Events in that gather are quite flat and laterally coherent, indicating that the picked velocity does a good job of performing NMO correction. The NMO stack, which is generated by applying the NMO correction corresponding to the lowest cost continuation velocity to all gathers and then summing over offset, is shown in Figure 5.15(a). Energy in the stack is both focused and laterally coherent, again indicating that the NMO correction that was applied has worked as intended. The diffraction data in Figure 5.15(b) appear as expected – most energy present in

that data have the hyperbolic moveout associated with seismic diffraction. Migration using the lowest cost continuation picked velocity produces a complete image in Figure 5.16(a) that has well defined, laterally coherent reflections that are often interrupted by discontinuities indicative of faulting. The diffraction features present in Figure 5.16(b) are collapsed to points, providing further confirmation of the quality of the picked velocities. These diffractions tend to delineate the faults which can be seen in the discontinuities of Figure 5.16(b).

The lowest cost final velocity model output by the continuation picking method produces quality complete and diffraction seismic images when used as part of a seismic processing workflow on the Gulf of Mexico field dataset featured in this section.

Heidrun Field Horizon Picking

To illustrate the versatility of the proposed approach I apply it to an automatic horizon picking problem. Seismic horizons are laterally continuous features in seismic images representing isochrons, or constant levels of geologic time (Vail, 1977). Identifying and mapping seismic horizons is a key step in seismic interpretation, enabling earth scientists to delineate subsurface structures, stratigraphy, and the volumetrics of potential subsurface reservoirs (Wu and Hale, 2013). Because this activity is such an essential step in subsurface evaluation, much research has focused on developing computer algorithms to automatically accomplish the task (Lomask et al., 2006; Fomel, 2010; Hoyes and Cheret, 2011; Wu and Hale, 2015; Wu and Fomel, 2018b; Xue et al., 2018; Peters et al., 2019). The approach presented here is not intended to replace this body of research, but rather to be seen as a tool that could work in tan-

dem with it. A popular approach for automatic horizon picking involves outputting a volume containing the likelihood for a horizon to exist at each point in space, for example Shi et al. (2020) use waveform embedding to extract seismic horizons using a deep convolutional network. The output of this approach is a horizon probability volume, and the horizon is determined by picking the large values. The variational picking approach employed presented here meshes well with such advanced techniques which output semblance-like volumes from which a surface is determined, and because smoothness is not explicitly imposed with the variational approach, it has the ability to track the rugose features frequently observed in seismic data.

As this is intended as a proof of concept of how the variational picking algorithm can be used to determine horizons, the semblance-like volume for horizon likelihood employed here is less advanced than some of those appearing in the literature. Instead, a trace representative of a seismic horizon is selected, cosine tapering is applied to its edges, and it is padded with zeros. This ideal horizon reference trace is called $h(t)$. To measure how well that ideal waveform matches with other traces throughout the volume, the cross correlation is calculated throughout the volume over shift τ,

$$\gamma(\tau, \mathbf{x}) = \int h(t+\tau) d(t, \mathbf{x}) d\tau, \tag{5.138}$$

where $d(t, \mathbf{x})$ is the seismic image. In order to ensure that the semblance-like panel non-negative, let

$$\alpha(\tau, \mathbf{x}) = \gamma(\tau, \mathbf{x}) - \min_{(\tau, \mathbf{x})} [\gamma(\tau, \mathbf{x})] \tag{5.139}$$

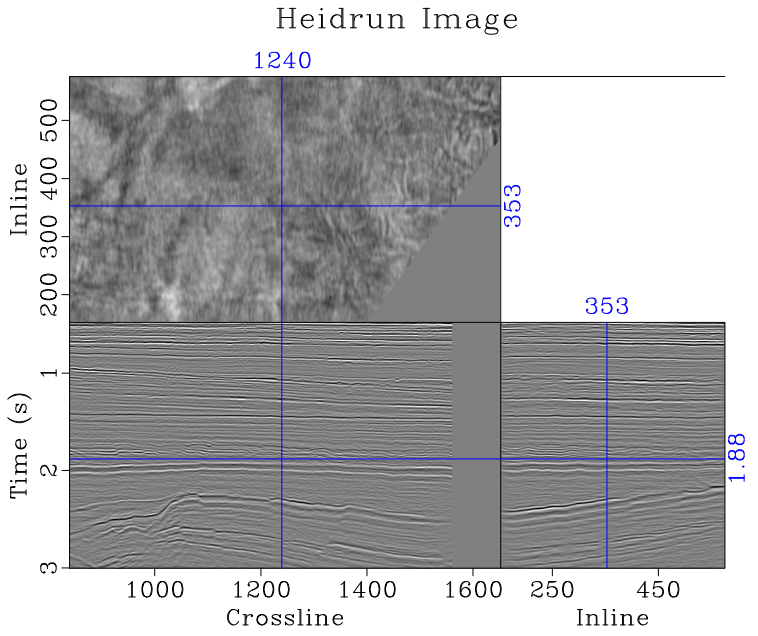
Automatic picking may then be performed on the correlation volume $\alpha(\tau, \mathbf{x})$ to determine the shifts defining the horizon, and those converted back to time in the image

using a by adding a reference time for the ideal horizon trace, taken here to be the time at that trace's midpoint.

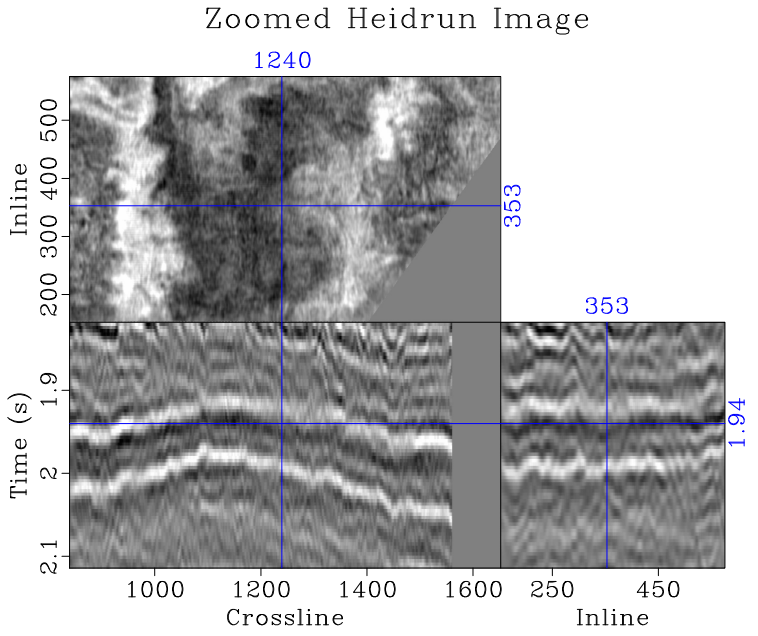
Figure 5.17(a) contains a seismic image from the Heidrun field off the coast of Norway, and Figure 5.17(b) contains a zoomed portion of that image. I will focus on automatically picking the horizon defined by the "double white peak" in Figure 5.17(b). A waveform representing an ideal version of that horizon is selected, and that horizon reference trace is plotted in Figure 5.18.

Cross correlation values between the horizon reference trace and the seismic image are performed over a variety of shifts according to Equations 5.138 and 5.139. The shift coordinate τ is transformed to image time t to be representative of horizon location and displayed in Figure 5.19. Continuation picking is applied to the cross correlation volume using a flat starting model of 1.9745 s using 12 continuation levels. The evolution of the picked horizon every ten iterations beginning with the starting model is shown in Figures 5.20(a) through 5.20(i). The final picked horizon is shown with constant horizon time contours in Figure 5.21(a), and without contours in Figure 5.21(b). The cost convergence is calculated for each model update using the semblance-like volume in Figure 5.19 and displayed in Figure 5.22.

To illustrate how the picked horizon looks in the seismic volume I generate a series of overlay images. Note that the position of the horizon is taken to be at the halfway point of the reference trace in Figure 5.18, and thus I would expect it to approximately track the trough between the double white peaks in the image. Constant inline slices are shown for inline 100 in Figure 5.23(a) and inline 300 in Fig-



(a)



(b)

Figure 5.17: (a) Seismic image from the Heidrun field; (b) zoomed portion of that image centered on the horizon that will be automatically picked.

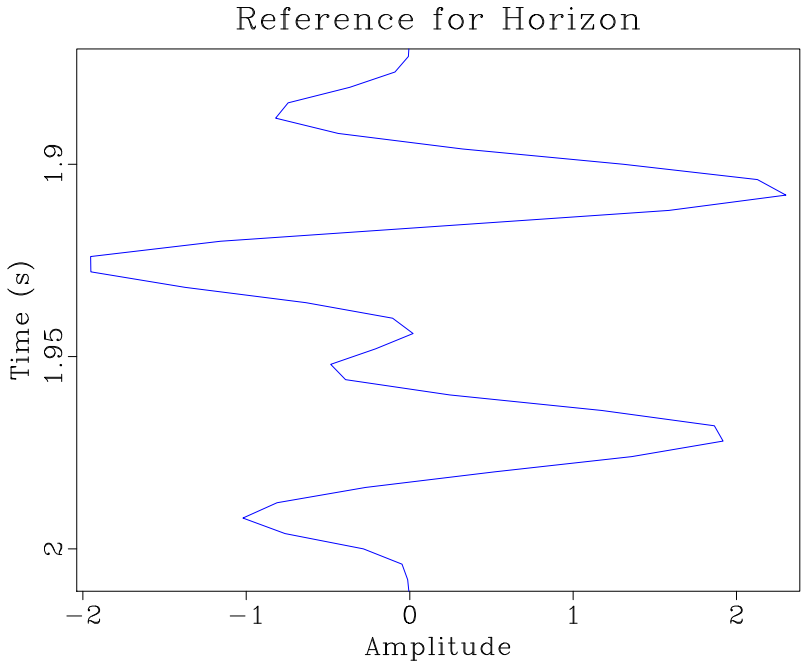


Figure 5.18: Reference trace for horizon that will be picked from Figure 5.17(b).

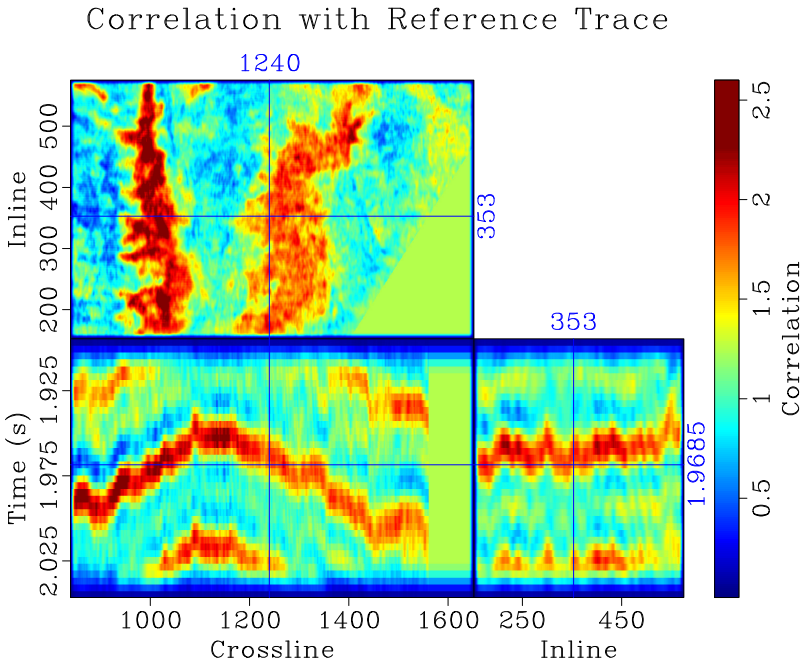


Figure 5.19: Cross correlation between reference trace and seismic image.

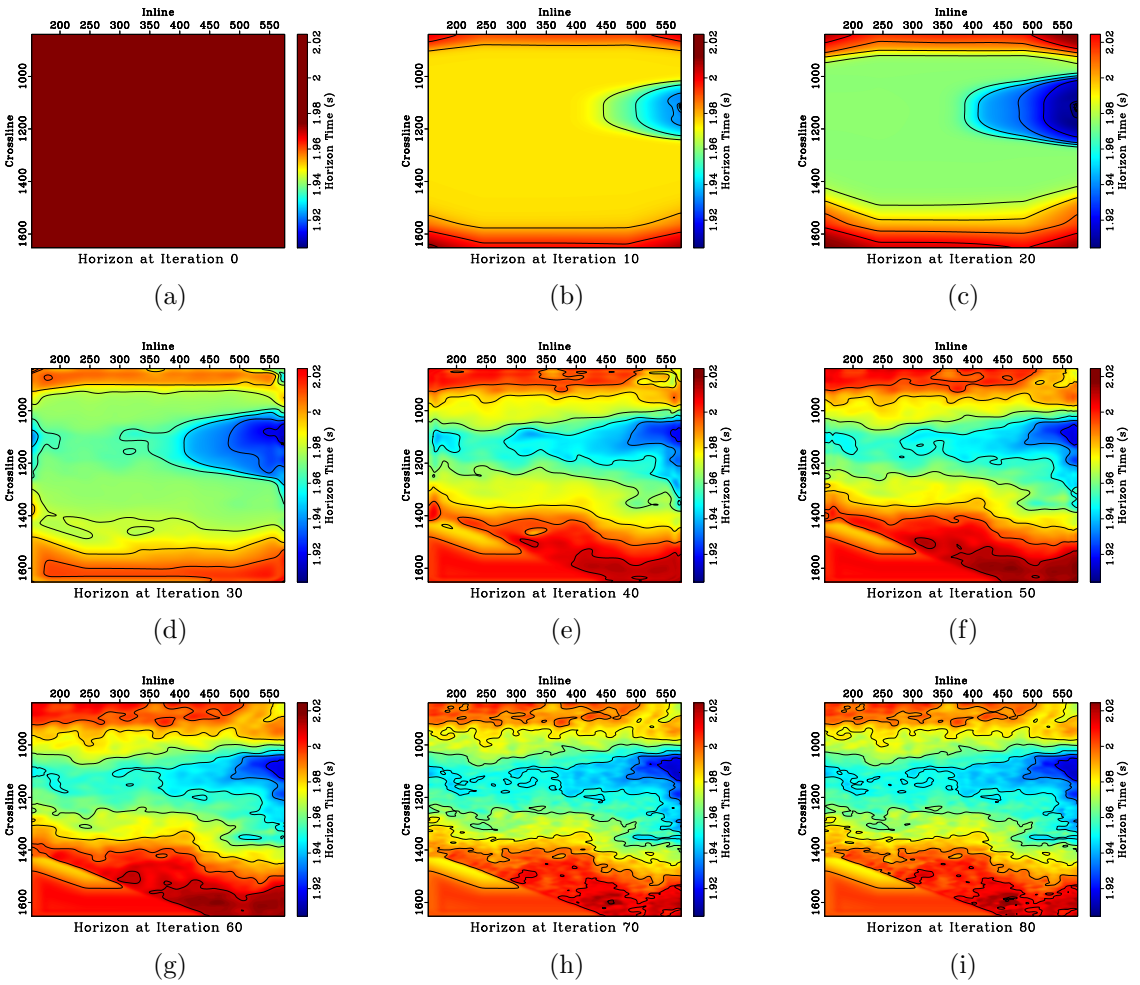
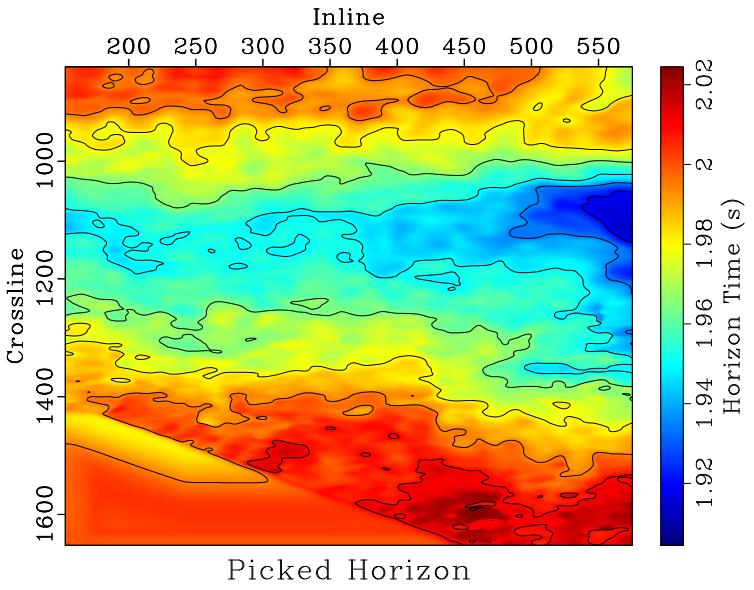
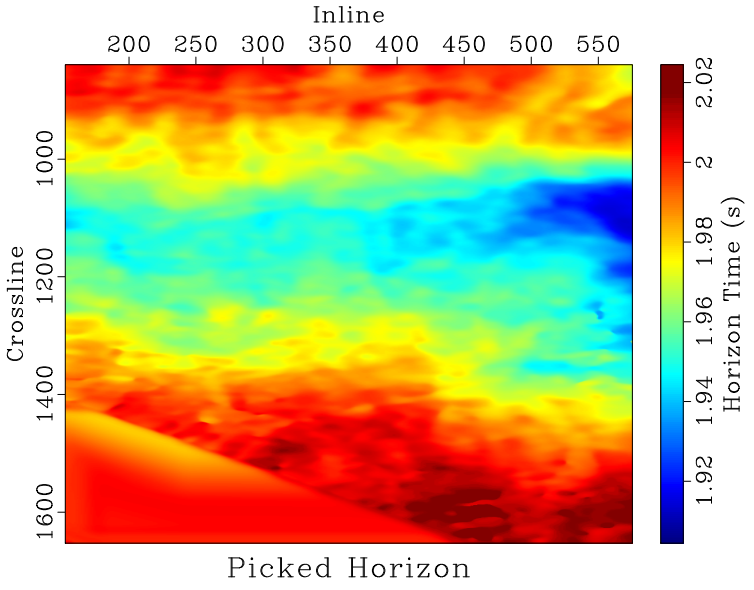


Figure 5.20: Evolution of continuation picking horizon overlaid by constant horizon time contours.



(a)



(b)

Figure 5.21: Final picked horizon (a) with constant horizon time contours; (b) without contours

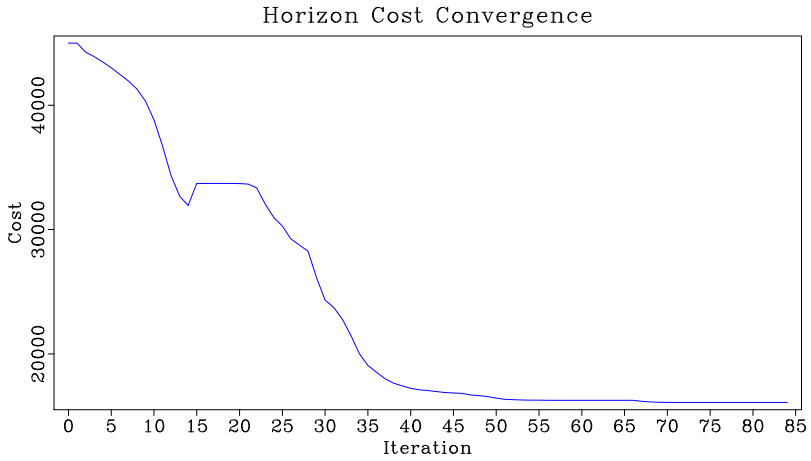


Figure 5.22: Cost convergence for continuation horizon picking using the semblance-like volume shown in Figure 5.19.

ure 5.23(b). Constant crossline slices are similarly generated and shown for crossline 1200 in Figure 5.23(c) and crossline 1400 in Figure 5.23(d).

The variational picking method outlined here appears to successfully pick the desired horizon from a poorly informed, flat, starting model. Examining the picked horizon image overlays, Figures 5.23(a) through 5.23(d), the horizon successfully tracks the trough between the two bright white peaks, even when that position changes relatively rapidly, as is the case near crossline 1050 in Figure 5.23(a), or when the character of the horizon wavelet changes somewhat, as is the case near inline 560 in Figure 5.23(d). The displays of the picked horizon in Figures 5.21(a) and 5.21(b) show how the method is able to identify a plunging anticline structure in the horizon. The horizon quality of fit volume defined by Equation 5.139 is primitive, and may not be well suited for the automatic interpretation of other horizons which are faulted or feature significant changes in wavelet. However, when applied to the horizon in

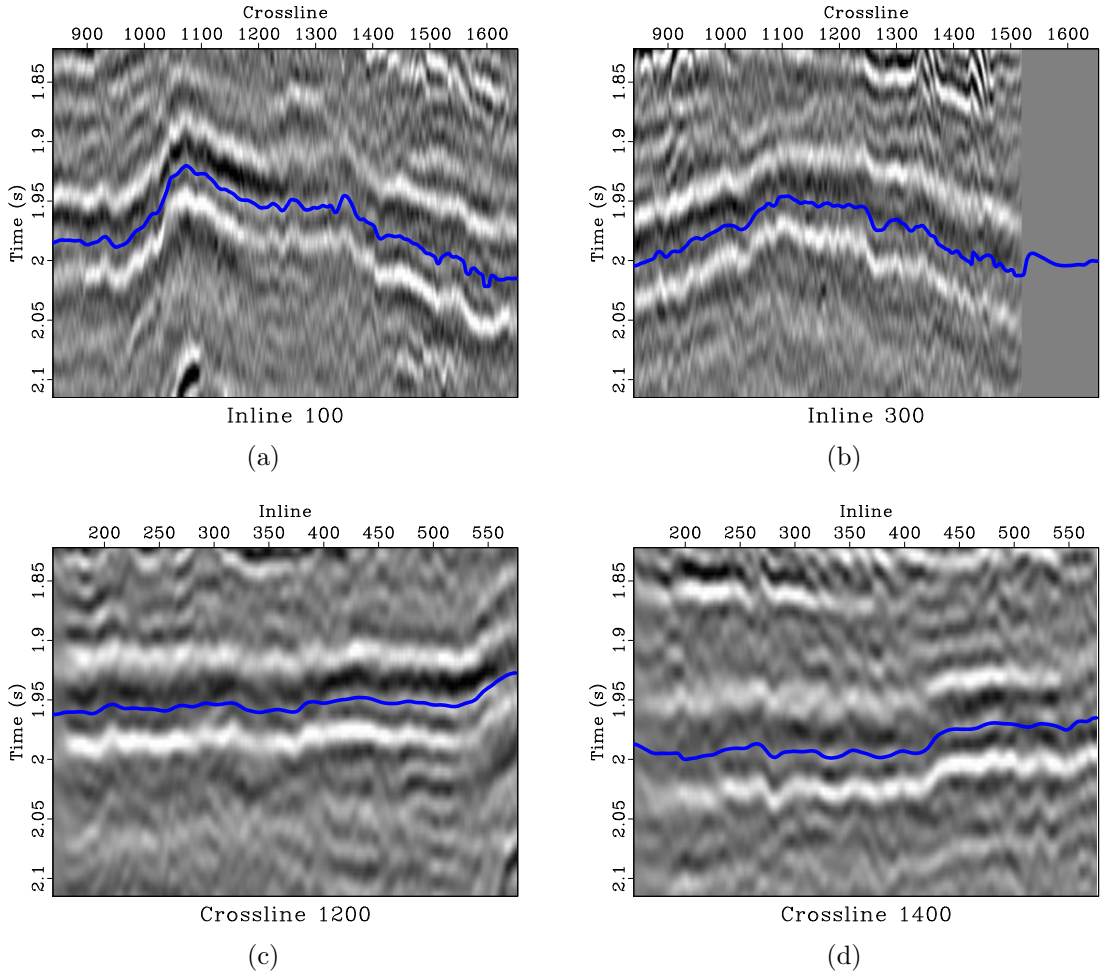


Figure 5.23: Picked horizon from Figure 5.21(a) overlaid on Heidrun seismic image from Figure 5.17(b) for: (a) Inline 100; (b) Inline 300; (c) Crossline 1200; (d) Crossline 1400.

this study and used in conjunction with the continuation variational picking method, it is is able to produce a quality representation of the horizon from the single seed wavelet shown in Figure 5.18. This result is promising, and indicates that when used in conjunction with more sophisticated measures of horizon location probability the approach outlined here could help produce high quality automatic interpretations of subsurface features with minimal human intervention.

CONCLUSIONS

I propose a variational method for picking velocity surfaces from semblance-like volumes that is guaranteed to converge to a minima in an infinite dimensional setting assuming that for each minima, there is a ball of non-zero radius in H^1 where it is the only critical point. When coupled with a continuation approach, this method is able to avoid many local minima, and when discretized and used on field data a ℓ-BFGS algorithm is able to help the method converge more rapidly.

Applying the method to a Viking Graben field dataset illustrates how when used in conjunction with continuation, the approach is able to determine geologically plausible velocity fields from DMO stack power volumes which can differ substantially from the starting model, and achieve lower costs than without utilizing continuation. Using the continuation velocity for DMO stacking and migration creates images with more laterally coherent reflectors, and performing Dix velocity analysis on the picked velocity reveals a well-defined velocity anomaly that the method is able to resolve.

I use a field dataset from the Gulf of Mexico to illustrate how when paired with

continuation, the proposed approach behaves more like a global minimizer. Applying the method with continuation to 125 constant gradient starting models leads to final models with similarly low costs which only differ substantially in areas where no reflection energy is present to guide the NMO scan the cost minimizing surface is picked from. Comparing the continuation result to one without continuation shows how continuation enables the method to evade numerous local minima. Using the lowest cost output model in a seismic processing workflow produces quality images, and further shows how the method may be incorporated as a tool for seismic processors.

The versatility of the method is demonstrated by using it to automatically pick a seismic horizon from the Heidrun field. Because smoothing is not explicitly imposed on the model using this method, but rather only used during the continuation process, the approach is able to determine a seismic horizon that tracks the image well and is able to change rapidly in space to follow the reflector.

The variational method outlined here for determining optimal surfaces from semblance-like volumes is more computationally expensive than existing methods for determining optimal lines from semblance panels, but the ability of the method to incorporate spatially adjacent information without explicitly imposing smoothing on the model justifies the added expense. Direct application of the method suffers from the multimodality and non-convexity of the proposed objective function, and unless a particularly well informed starting model is used the method is likely to converge to a local minima. Such local minima may differ significantly from the global minimum. This difficulty may be overcome by applying the variational picking scheme with a continuation approach. Although continuation increases the cost of the method

through the creation of additional smoothed semblance volumes and applying the picking scheme to each smoothed volume, the ability of the approach to avoid local minima and find a superior final model justifies the expense.

Beyond the demonstrated applications shown here for processing 2D seismic lines and automatically interpreting seismic horizons, this variational picking method may be used to generate starting models for full waveform inversion, extended to picking 3D velocity volumes from 4D semblance hypervolumes, or applied to other situations where one wishes to determine a laterally continuous surface, such as calculating time shifts to match time-lapse seismic volumes.

Chapter 6

Conclusion

This dissertation explores several techniques for selecting parameters used in seismic processing applications. These parameters are often tied to seismic wave propagation, which itself is closely related to properties of the propagating medium. Therefore, accurately determining these parameters can provide insight into the subsurface materials as well as enabling the creation of subsurface images that are better resolved, more accurate, and less distorted.

In Chapter 2 I propose an efficient algorithm that corrects for the residual effects of anisotropy in seismic images. Applying this algorithm to a field dataset creates higher resolution, more coherent seismic images featuring increased bandwidth. The anisotropic azimuth and intensity information it generates provides insight into the posible location and orientation of subsurface fracture networks and the dominant stress field, which is valuable for predicting and planning reservoir production.

In Chapter 3 I introduce oriented velocity continuation, which generates a suite of slope-decomposed seismic images over a range of seismic migration velocities by transporting slope-decomposed images over their characteristics. I show how oriented velocity continuation can be used in the context of seismic diffraction imaging to perform migration velocity analysis by selecting the migration velocity that leads to

the most coherent energy in slope gathers.

Chapter 4 uses the tool of oriented velocity continuation in conjunction with the concept of path-integral imaging to create probabilistic diffraction images. These images are generated by using weights tied to the likelihood of a correctly imaged diffraction existing at a location within an image given a migration velocity. The probabilistic imaging process creates diffraction images featuring reduced noise, and the weighting process suppresses wave field components not corresponding to diffraction. This method works complementarily to data domain diffraction separation techniques, and the imaging process outputs the most likely, or expectation, seismic migration velocity at each subsurface position as well as a measure of confidence in that velocity. Applying the probabilistic method to the same field dataset used in Chapter 3 yields a more detailed, more geologically plausible seismic velocity field than the method used in that section. The expectation velocity is able to resolve interesting features like a seismic velocity inversion. The corresponding seismic diffraction image clearly displays diffractions corresponding to the seafloor, thrust faults, the tectonic plate boundary, and the transition from sedimentary to crystalline rock.

In Chapter 5 I present a variational method for determining best fit surfaces from semblance-like volumes by minimizing an associated functional. This method is able to find the surface that best represents the evolution of a seismic parameter, like seismic velocity, through a volume measuring the quality of fit for each parameter value at each location. I prove that minimizers to the functional exist, and that a gradient descent scheme will converge to those minimizers in an infinite dimensional setting. I use a ℓ-BFGS algorithm to accelerate that convergence for finite dimen-

sional implementation. Employing continuation, or graduated optimization, I show how the method is able to determine geologically plausible velocity surfaces that are largely independant of the starting model and avoid many of the local minima the picking functional possesses. The velocity fields found by the method are used in seismic processing workflows, illustrating how it can be a valuable tool for seismic data processing and imaging. The versatility of the approach is also demonstrated by using it for automatic interpretation of a seismic horizon in a 3D image.

In summary, the primary research contributions of this dissertation incude:

1. Developing an algorithm based on dynamic programming to efficiently correct for the effects of anisotropy in seismic depth images, determine the fastest axis of wave propagation, and provide a measure of the relative difference in velocity between the fastest and slowest wave propagation orientation.

2. Presenting a method for efficiently creating a collection of seismic images over a range of seismic velocities by transporting slope decomposed images along their characteristics, and illustrating how this method may be used in conjunction with seismic diffraction imaging to determine migration velocity with limited offset data.

3. Showing how a collection of slope decomposed seismic diffraction images created using a range of velocities may be used in conjunction with path-integral imaging to highlight features with a high likelihood of being seismic diffractions, suppress noise and other signal which is not likely related to diffraction,

and automatically output the most likely seismic migration velocity as well as a measure of confidence in that velocity.

4. Proposing a variational method which is able to determine the best fit parameter surface from a volume measuring parameter fit quality.

FUTURE WORK

The research in this dissertation touches on many topics, leading to numerous promising avenues for future inquiry.

Determining a mathematical relationship between the anisotropic azimuth and intensity found in Chapter 2 and various anisotropic parameters commonly used in seismic data processing could enable the method to be utilized as an intermediate step in anisotropic seismic processing. In such a situation it would be able to determine parameters related to seismic anisotropy relatively inexpensively so they may be used in more accurate and more expensive operations. The proposed approach in Chapter 2 also measures anisotropy along a complete ray path, so finding an accurate way to convert that to a local anisotropy attribute would be useful in situations where seismic anisotropy can change relatively rapidly in space.

The equations shown in Chapters 3 and 4 for oriented velocity continuation and probabilistic diffraction imaging are valid for 3D imaging even though only 2D examples were performed. Expanding the oriented velocity continuation framework to 3D could lead to superior results, as it would be able to accommodate the effects of out of plane diffractions. This would require developing a tool for decomposing 3D

seismic data into 5D dimensions, with two added dimensions for slope. Such a process would require a large amount of memory, making it an interesting high performance computing challenge. Once data are separated in their constituent slope components, parallelizing the velocity continuation process in the Fourier domain should be relatively straightforward. Another promising direction for inquiry is developing continuation operators representing different aspects of seismic wave propagation. These could also be incorporated in a probabilistic imaging framework, although adding additional parameters could dramatically increase the memory requirements for the operation. Exploring other probabilistic weights, or finding better measures of diffraction likelihood than semblance which are able to recognize the polarity reversals in edge diffractions could also improve the method.

The formulation for variational picking provided in Chapter 5 is valid for picking a parameter volume from a semblance-like hypervolume, which would likely be a valuable extension of the method. The memory requirements for this expansion could be quite substantial, so thoughtful implementation in a high performance setting would be needed. Using a constrained optimization framework to enforce apriori information about the parameter being chosen, such as the reasonable limits on interval velocities within an area, could help the method from producing surfaces that track anomalous peaks in the semblance-like volume that do not correspond to high parameter quality. Applying the method to semblance-like volumes made with other attributes like differential semblance (Symes, 1999), which could reduce the need for continuation as it possesses a global minimum, or selective correlation (Larner and Celis, 2007), which could increase resolution, could also prove useful. Exploring

different methods of continuation or graduated optimization could reduce the computational expense of the picking process and enable it to converge more rapidly while avoiding numerous local minima. Finally, using the variational picking method in conjunction with sophisticated techniques used for determining how the probability of a horizon's position evolves through space, like those based on convolutional neural networks and deep learning, could enable high quality automatic interpretation of seismic features, freeing interpreters to spend time on more challenging aspects of subsurface evaluation.

The concepts explored in this thesis enable seismic processors in the selection of parameters to use in their workflows, facilitating the creation of accurate, high quality representations of the subsurface. These tools can help Earth scientists safely and economically produce resources that will fill the growing demand for energy that fuels the world's economy, hopefully continuing the last century's trend of reducing poverty, increasing wealth, and improving life expectancy for all humanity at an unprecedented rate.

The experiments using non-proprietary data in this dissertation are reproducible in the `Madagascar` open source software environment (Fomel et al., 2013), an open source multidimensional data analysis software library featuring all the code used in this work. To download the software, please visit `https://www.reproducibility.org` or `https://www.ahay.org`.

Bibliography

Abedi, M. M., A. Stovas, and Y. Ivanov, 2019, Acoustic wave propagation in orthorhombic media: Phase velocity, group velocity, and moveout approximations: Geophysics, **84**, no. 6, C269–C279.

Alder, F., and S. Brandwood, 1999, Robust estimation of dense 3-D stacking velocities from automated picking: 69th Annual International Meeting: Society of Exploration Geophysicists Expanded Abstracts, 1162–1165.

Alkhalifah, T., and K. Larner, 1994, Migration error in transversely isotropic media: Geophysics, **59**, no. 9, 1405–1418.

Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, no. 5, 1550–1566.

Alkhalifah, T., I. Tsvankin, K. Larner, and J. Toldi, 1996, Velocity analysis and imaging in transversely isotropic media: Methodology and a case study: The Leading Edge, **15**, no. 5, 371–378.

Armijo, L., 1966, Minimization of functions having Lipschitz continuous first partial derivatives.: Pacific Journal of Mathematics, **16**, 1 – 3.

Arnaud, J. D., J. P. Dunand, and V. Curinier, 2004, High density picking for accurate velocity and anisotropy determination: 74th Annual International Meeting: Society of Exploration Geophysicists Expanded Abstracts, 1627–1629.

Babich, V. M., and V. S. Buldyrev, 1972, Short-Wavelength Diffraction Theory

Asymptotic Methods, 1 ed.: Springer-Verlag, volume **4** *of* Springer Series on Wave Phenomena.

Bader, S., X. Wu, and S. Fomel, 2019, Missing log data interpolation and semiautomatic seismic well ties using data matching techniques: Interpretation, **7**, T347–T361.

Bangs, N. L. B., G. F. Moore, S. P. S. Gulick, E. M. Pangborn, H. J. Tobin, S. Kuramoto, and A. Taira, 2009, Broad, weak regions of the Nankai Megathrust and implications for shallow coseismic slip: Earth and Planetary Science Letters, **284**, no. 1, 44–49.

Bashkardin, V., T. J. Browaeys, S. Fomel, F. Gao, S. A. Morton, S. Terentyev, and A. Vladimirsky, 2012, Phase-space computation of multi-arrival traveltimes: Part II – Implementation and application to angle-domain imaging: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–6.

Berger, M. S., 1977, Nonlinearity and functional analysis: lectures on nonlinear problems in mathematical analysis: Academic press, **74**.

Berkovitch, A., I. Belfer, Y. Hassin, and E. Landa, 2009, Diffraction imaging by multifocusing: Geophysics, **74**, no. 6, WCA75–WCA81.

Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation methods: Geophysics, **69**, no. 5, 1283–1298.

Blake, A., and A. Zisserman, 1987, Visual reconstruction.

Born, M., and E. Wolf, 1959, Principles of Optics: Pergamon Press Inc.

BP, 2020, Statistical Review of World Energy 2020.

Brandsberg-Dahl, S., B. Ursin, and M. V. de Hoop, 2003, Seismic velocity analysis in the scattering-angle/azimuth domain: Geophysical Prospecting, **51**, 295–314.

Brown, A. R., 2011, Interpretation of Three-Dimensional Seismic Data: American Association of Petroleum Geologists.

Burnett, W., and S. Fomel, 2009, 3D velocity-independent elliptically anisotropic moveout correction: Geophysics, **74**, no. 5, WBI129–WBI136.

Burnett, W., and S. Fomel, 2011, Azimuthally anisotropic 3D velocity continuation: International Journal of Geophysics, Article ID 484653.

Burnett, W., S. Fomel, and R. Bansal, 2011, Diffraction velocity analysis by path-integral seismic imaging: 81st Annual International Meeting, SEG, Expanded Abstracts, 3898–3902.

Byrd, R. H., and R. A. Tapia, 1975, An extension of Curry's theorem to steepest descent in normed linear spaces: Mathematical Programming, **9**, 247–254.

Candès, E., L. Demanet, and L. Ying, 2009, A fast butterfly algorithm for the computation of Fourier integral operators: Multiscale Modeling & Simulation, **7**, 1727–1750.

Chambolle, A., and P.-L. Lions, 1997, Image recovery via total variation minimization and related problems: Numerische Mathematik, **76**, 167–188.

Chapelle, O., M. Chi, and A. Zien, 2006, A Continuation Method for Semi-Supervised SVMs: Proceedings of the 23rd International Conference on Machine Learning, Association for Computing Machinery, 185–192.

Chaudhuri, S., and A. Solar-Lezama, 2011, Smoothing a program soundly and robustly: Proceedings of the 23rd International Conference on Computer Aided Ver-

ification, Springer-Verlag, 277–292.

Cheng, J., J. Geng, H. Wang, and Z. Ma, 2011, 3D Kirchhoff prestack time migration in average illumination-azimuth and incident-angle domain for isotropic and vertical transversely isotropic media: Geophysics, **76**, no. 1, S15–S27.

Claerbout, J., 1993, Earth soundings analysis: Processing versus inversion: Stanford University Press. Stanford Exploration Project.

Claerbout, J., 2008, Basic earth imaging: Stanford University.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Claerbout, J. F., 1985, Imaging the earth's interior: Blackwell Scientific Publications, Ltd.

Corrigan, D., R. Withers, J. Darnall, and T. Skopinski, 1996, Fracture mapping from azimuthal velocity analysis using 3-D surface seismic data: 66th Annual International Meeting, Society of Exploration Geophysicists Expanded Abstracts, 1834–1837.

Courant, R., and D. Hilbert, 1989, Methods of mathematical physics: John Wiley & Sons.

Crampin, S., 1981, Review of wave motion in anisotropic and cracked elastic-media: Wave Motion, **3**, no. 4, 343–391.

Crampin, S., 1984a, Anisotropy in exploration seismics: First Break, **2**, no. 3, 19–21.

Crampin, S., 1984b, An introduction to wave propagation in anisotropic media: Geophysical Journal of the Royal Astronomical Society, **76**, no. 1, 17–28.

Crampin, S., 1985, Evidence for aligned cracks in the earth's crust: First Break, **3**, no. 3, 16–20.

Crampin, S., 1986, Anisotropy and transverse isotropy: Geophysical Prospecting, **34**, no. 1, 94–99.

Curry, H. B., 1944, The method of steepest descent for non-linear minimization problems: Quarterly of Applied Mathematics, **2**, 258–261.

Dacorogna, B., 2004, Introduction to the calculus of variations: Imperial College Press. EBL-Schweitzer.

de Bazelaire, E., 1988, Normal moveout revisited: Inhomogeneous media and curved interfaces: Geophysics, **53**, 143–157.

de Figueiredo, J. J. S., F. Oliveira, E. Esmi, L. Freitas, J. Schleicher, A. Novais, P. Sussner, and S. Green, 2013, Automatic detection and imaging of diffraction points using pattern recognition: Geophysical Prospecting, **61**, 368–379.

Decker, L., 2014, Seismic diffraction imaging methods and applications: M.S. Thesis: The University of Texas at Austin.

Decker, L., and S. Fomel, 2014, Diffraction imaging and velocity analysis using oriented velocity continuation: 84th Annual International Meeting, SEG, Expanded Abstracts, 4810–4815.

Decker, L., and S. Fomel, 2018, A finite-element method for blind deconvolution with dynamic frequency wavelets: SEG Technical Program Expanded Abstracts 2018, 4563–4567.

Decker, L., and S. Fomel, 2019, Path-integral seismic diffraction imaging with probability weights: SEG International Exposition and Annual Meeting, 4231–4235.

Decker, L., and S. Fomel, 2020, A variational method for picking velocity surfaces from semblance scans: SEG Technical Program Expanded Abstracts 2020, 3684–3688.

Decker, L., and S. Fomel, 2021a, A continuation approach for avoiding local minima in seismic velocity picking: SEG Technical Program Expanded Abstracts 2021, Submitted.

Decker, L., and S. Fomel, 2021b, A probabilistic approach to seismic diffraction imaging: Lithosphere, In Review.

Decker, L., and S. Fomel, 2021c, A variational approach for picking optimal surfaces from semblance-like panels using continuation: Geophysics, Submitted.

Decker, L., X. Janson, and S. Fomel, 2015, Carbonate reservoir characterization using seismic diffraction imaging: Interpretation, **3**, no. 1, SF21–SF30.

Decker, L., and A. Klokov, 2014, Diffraction extraction by plane-wave destruction of partial images: 84th Annual International Meeting, SEG, Expanded Abstracts, 3862–3867.

Decker, L., A. Klokov, and S. Fomel, 2013, Comparison of seismic diffraction imaging techniques: plane wave destruction versus apex destruction: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4054–4059.

Decker, L., D. Merzlikin, and S. Fomel, 2017a, Diffraction imaging and time-migration velocity analysis using oriented velocity continuation: Geophysics, **82**, no. 2, U25–U35.

Decker, L., D. Merzlikin, and S. Fomel, 2017b, Enhancing seismic-diffraction images using semblance-weighted least-squares migration: 87th Annual International

Meeting, SEG, Expanded Abstracts, 5294–5299.

Decker, L., and Q. Zhang, 2019, Correcting residual HTI moveout and determining principal anisotropic azimuth in arbitrarily sampled image gathers using dynamic time warping: SEG Technical Program Expanded Abstracts 2019, 404–408.

Decker, L., and Q. Zhang, 2020, Quantifying and correcting residual azimuthal anisotropic moveout in image gathers using dynamic time warping: Geophysics, **85**, O71–O82.

Dennis, Jr., J. E., and J. J. Moré, 1977, Quasi-Newton methods, motivation and theory: SIAM Review, **19**, 46–89.

Deregowski, S., 1986, What is DMO?: First Break, **4**.

Deregowski, S., 1990, Common-offset migrations and velocity analysis: First Break, **8**.

Deschamps, T., and L. D. Cohen, 2001, Fast extraction of minimal paths in 3D images and applications to virtual endoscopy: Medical image analysis, **5**, 281–99.

Dobson, D. C., and C. R. Vogel, 1997, Convergence of an Iterative Method for Total Variation Denoising: SIAM Journal on Numerical Analysis, **34**, 1779–1791.

Doicin, D., C. Johnson, N. Hargreaves, and C. Perkins, 1994, Machine–guided velocity interpretation: SEG Technical Program Expanded Abstracts, 1413–1416.

Engquist, B., and O. Runborg, 2003, Computational high frequency wave propagation: Acta Numerica, **12**, 181–266.

Evans, L. C., 2010, Partial differential equations, 2nd ed.: American Mathematical Society.

Feenstra, R. C., R. Inklaar, and M. P. Timmer, 2015, The Next Generation of the

Penn World Table: American Economic Review, **105**, no. 10, 3150–3182. (available for download at https://www.ggdc.net/pwt).

Fomel, S., 2002, Applications of plane wave destruction filters: Geophysics, **67**, no. 6, 1946–1960.

Fomel, S., 2003a, Angle-domain seismic imaging and the oriented wave equation: 73rd Annual International Meeting, SEG, Expanded Abstracts, 893–898.

Fomel, S., 2003b, Time migration velocity analysis by velocity continuation: Geophysics, **68**, no. 5, 1662–1672.

Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, no. 3, A29–A33.

Fomel, S., 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

Fomel, S., 2009a, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, V25–V33.

Fomel, S., 2009b, Velocity analysis using AB semblance: Geophysical Prospecting, **57**, 311–321.

Fomel, S., 2010, Predictive painting of 3d seismic volumes: Geophysics, **75**, A25–A30.

Fomel, S., and H. Kaur, 2021, Wave-equation time migration: Geophysics, **86**, S103–S111.

Fomel, S., and E. Landa, 2014a, Structural uncertainty of time-migrated seismic images: Journal of Applied Geophysics, **101**, 27–30.

Fomel, S., E. Landa, and M. T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, no. 6, U89–U94.

Fomel, S., and M. Prucha, 1999, Angle-gather time migration: SEP-100, Stanford Exploration Project, 141–150.

Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar: open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, **1**, no. 1, e8.

Fomel, S. B., and E. Landa, 2014b, Structural uncertainty of time-migrated seismic images: Journal of Applied Geophysics, **101**, 27–30.

Forel, D., T. Benz, and W. D. Pennington, 2005, Seismic Data Processing with Seismic Un*x: A 2D Seismic Data Processing Primer: Society of Exploration Geophysicists.

Fowler, P., 1988, Seismic velocity estimation using prestack time migration: PhD thesis, Stanford University.

Gallego, F., J. J. Quintero, and J. C. Riaño, 2015, Convergence of the steepest descent method with line searches and uniformly convex objective in reflexive Banach spaces: Mathematical Communications, **20**, 161–173.

Gazdag, J., 1978, Wave equation migration with the phase-shift method: Geophysics, **43**, 1342–1351.

Gelfand, I. M., and S. V. Fomin, 2000, Calculus of variations: Dover Publications, Inc.

Ghosh, S., and S. Fomel, 2012, Multiple suppression in the t-x-p domain: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–6.

Grechka, V., 2009, Applications of seismic anisotropy in the oil and gas industry: EAGE Publications.

Grechka, V., A. Pech, and I. Tsvankin, 2005, Parameter estimation in orthorhombic

media using multicomponent wide-azimuth reflection data: Geophysics, **70**, no. 2, D1–D8.

Grechka, V., and I. Tsvankin, 1998, 3-D description of normal moveout in anisotropic inhomogenous media: Geophysics, **63**, no. 3, 1079–1092.

Greenberg, M. D., 1978, Foundations of Applied Mathematics: Dover Publications, Inc.

Hale, D., 1984, Dip-moveout by Fourier transform: Geophysics, **49**, 741–757.

Hale, D., 2013, Dynamic warping of seismic images: Geophysics, **78**, no. 2, S105–S115.

Harlan, W., J. Claerbout, and F. Rocca, 1984, Signal to noise separation and velocity estimation: Geophysics, **49**, no. 11, 1869–1880.

Harlan, W. S., 2001, Constrained automatic moveout picking from semblances: http://billharlan.com/pub/papers/autopick.pdf.

Hazan, E., K. Y. Levy, and S. Shalev-Shwartz, 2016, On Graduated Optimization for Stochastic Non-Convex Problems: Proceedings of The 33rd International Conference on Machine Learning, PMLR, 1833–1841.

Helbig, K., 1994, Foundations of elastic anisotropy for exploration seismics: Pergamon Press.

Helbig, K., and L. Thomsen, 2005, 75-plus years of anisotropy in exploration and reservoir seismics: A historical review of concepts and methods: Geophysics, **70**, no. 6, 2–32.

Hill, N. R., 1990, Gaussian beam migration: Geophysics, **55**, 1416–1428.

Hoyes, J., and T. Cheret, 2011, A review of "global" interpretation methods for automated 3d horizon picking: The Leading Edge, **30**, 38–47.

Hu, J., S. Fomel, and L. Ying, 2015, A fast algorithm for 3D azimuthally anisotropic velocity scan: Geophysical Prospecting, **63**, 368–377.

Hubral, P., and T. Krey, 1980, Interval velocities from seismic reflection time measurements: Society of Exploration Geophysicists.

Iserles, A., 1996, A First Course in the Numerical Analysis of Differential Equations: Cambridge University Press.

Izmailov, A. F., and A. A. Tret'yakov, 1999, On the gradient method in a hilbert space in the case of nonisolated minima: Computational mathematics and mathematical physics, **39**, 521–524.

Karátson, J., 1999, Gradient method for non-injective operators in hilbert space with application to neumann problems: Applicationes Mathematicae, **26**, 333–346.

Karátson, J., et al., 2000, Gradient method for non-uniformly convex functionals in hilbert space: Pure Mathematics and Applications, **11**, 309–316.

Kaur, H., S. Fomel, and N. Pham, 2019, *in* Elastic wave-mode separation in heterogeneous anisotropic media using deep learning: 2654–2658.

Keller, J. B., 1962, Geometrical Theory of Diffraction: Journal of the Optical Society of America, **52**, 116–130.

Khaidukov, V., E. Landa, and T. Moser, 2004, Diffraction imaging by focusing-defocusing: an outlook on seismic super resolution: Geophysics, **56**, 1478–1490.

Kiefer, J., 1953, Sequential Minimax Search for a Maximum: Proceedings of the American Mathematical Society, **4**, 502–506.

Klem-Musatov, K., 1994, Theory of seismic diffractions: Society of Exploration Geophysicists.

Klem-Musatov, K., A. M. Aizenberg, J. Pajchel, and H. B. Helle, 2008, Edge and Tip Diffractions: Theory and Applications in Seismic Prospecting: Society of Exploration Geophysicists.

Klokov, A., and S. Fomel, 2012, Separation and imaging of seismic diffractions using migrated dip-angle gathers: Geophysics, **77**, no. 6, S131–S143.

Klokov, A., and S. Fomel, 2013, Selecting an optimal aperture in Kirchhoff migration using dip-angle gathers: Geophysics, **76**, no. 6, S243–S254.

Koren, Z., and I. Ravve, 2011, Full-azimuth subsurface angle domain wavefield decomposition and imaging. Part I: Directional and reflection image gathers: Geophysics, **76**, no. 1, S1–S13.

Kozlov, E., N. Baransky, E. Korolev, A. Antonenko, and E. Koshchuck, 2004, Imaging scattering objects masked by specular reflections: 74th Annual International Meeting, SEG, Expanded Abstracts, 1131–1134.

Lanczos, C., 1966, The variational principles of mechanics: University of Toronto Press.

Landa, E., 2012, Seismic diffraction: where's the value?: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–4.

Landa, E., S. Fomel, and T. Moser, 2006, Path-integral seismic imaging: Geophysical Prospecting, **54**, 491–503.

Landa, E., S. Fomel, and M. Reshef, 2008, Separation, imaging, and velocity analysis of seismic diffractions using migrated dip-angle gathers: 78th Annual International

Meeting, SEG, Expanded Abstracts, **27**, no. 1, 2176–2180.

Larner, K., and C. Beasley, 1987, Cascaded migrations - Improving the accuracy of finite-difference migration: Geophysics, **52**, no. 5, 618–643.

Larner, K., and V. Celis, 2007, Selective-correlation velocity analysis: Geophysics, **72**, U11–U19.

Lee, J. D., M. Simchowitz, M. I. Jordan, and B. Recht, 2016, Gradient descent only converges to minimizers: 29th Annual Conference on Learning Theory, PMLR, 1246–1257.

Li, D.-H., and M. Fukushima, 2001, A modified BFGS method and its global convergence in nonconvex minimization: Journal of Computational and Applied Mathematics, **129**, 15–35. (Nonlinear Programming and Variational Inequalities).

Li, J., and W. W. Symes, 2007, Interval velocity estimation via NMO–based differential semblance: Geophysics, **72**, U75–U88.

Liu, D. C., and J. Nocedal, 1989, On the limited memory BFGS method for large scale optimization: Math. Program., **45**, 503–528.

Liu, Y., and S. Fomel, 2013, Seismic data analysis using local time-frequency decomposition: Geophysical Prospecting, **74**, no. 3, 516–525.

Ljubič, J. I., and G. Maistrovskif, 1970, A general theory of relaxation processes for convex functionals, uspehi mat. nauk 25 (1970), no. 1 (151), 57–112: English transl. in Russian Math.

Lomask, J., A. Guitton, S. Fomel, J. Claerbout, and A. A. Valenciano, 2006, Flattening without picking: Geophysics, **71**, P13–P20.

Luenberger, D. G., and Y. Ye, 1984, Linear and nonlinear programming: Springer, **2**.

Mallick, S., K. L. Craft, L. J. Meister, and R. E. Chambers, 1997, Computation of principal directions of azimuthal anisotropy from P-wave seismic data: Exploration Geophysics, **28**, no. 4, 379–382.

Mavko, G., T. Mukerji, and J. Dvorkin, 2020, The rock physics handbook, 3 ed.: Cambridge University Press.

Mawhin, J., and M. Willem, 2010, Origin and evolution of the palais–smale condition in critical point theory: Journal of Fixed Point Theory and Applications, **7**, 265–290.

Merzlikin, D., and S. Fomel, 2017, Analytical path-summation imaging of seismic diffractions: Geophysics, **82**, no. 1, S51–S59.

Merzlikin, D., S. Fomel, and M. K. Sen, 2019, Least-squares path-summation diffraction imaging using sparsity constraints: Geophysics, **84**, S187–S200.

Merzlikin, D., S. Fomel, and X. Wu, 2020, Least-squares diffraction imaging using shaping regularization by anisotropic smoothing: Geophysics, **85**, S313–S325.

Mobahi, H., and J. W. Fisher, 2015, A theoretical analysis of optimization by gaussian continuation: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI Press, 1205–1211.

Moore, G., and T. Shipley, 1993, Character of the décollement in the Leg 131 area, Nankai Trough: ed Hill, IA, Proc. scientific results, ODP, Leg 131, Nankai Trough, 73–82.

Moore, G. F., N. L. Bangs, A. Taira, S. Kuramoto, E. Pangborn, and H. J. Tobin, 2007, Three-Dimensional Splay Fault Geometry and Implications for Tsunami Generation: Science, **318**, no. 5853, 1128–1131.

Moore, G. F., J. O. Park, N. L. Bangs, S. P. Gulick, H. J. Tobin, Y. Nakamura, S. Sato, T. Tsuji, T. Yoro, H. Tanaka, S. Urakai, Y. Kido, Y. Sanada, S. Kuramoto, A. Taira, and the Expedition 314/315/316 Scientists, 2009, Structural and seismic stratigraphic framework of the NanTroSEIZE Stage 1 transect: Proceedings of the Integrated Ocean Drilling Program, 1–46.

Moore, G. F., T. Shipley, P. Stoffa, D. Karig, A. Taira, S. Kuramoto, H. Tokuyama, and K. Suyehiro, 1990, Structure of the Nankai Trough accretionary zone from multichannel seismic reflection data: Journal of Geophysical Research: Solid Earth, **95**, no. B6, 8753–8765.

Mordecai, A., and D. J. Wilde, 1966, Optimality proof for the symmetric Fibonacci search technique: Fibonacci Quarterly, **4**, 265–269.

Moser, T., and C. Howard, 2008, Diffraction imaging in depth: Geophysical Prospecting, **56**, 627–641.

Mulder, W. A., and A. P. E. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, no. 1, 208–221.

Newendorp, P. D., 1976, Decision analysis for petroleum exploration.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, no. 151, 773–782.

OPEC, 2020, World oil outlook 2045: Organization of the Petroleum Exporting Countries.

Ottolini, R., 1983, Signal/Noise Separation in Dip Space: SEP-37, Stanford Exploration Project, 143–150.

Palais, R. S., and S. Smale, 1964, A generalized Morse theory: Bulletin of the American Mathematical Society, **70**, 165 – 172.

Penot, J.-P., 2002, On the convergence of descent algorithms: Comput. Optim. Appl., **23**, 279–284.

Peters, B., J. Granek, and E. Haber, 2019, Multiresolution neural networks for tracking seismic horizons from few training images: Interpretation, **7**, SE201–SE213.

Pham, N., S. Fomel, and D. Dunlap, 2019, Automatic channel detection using deep learning: Interpretation, **7**, SE43–SE50.

Polyak, B., 1963, Gradient methods for the minimisation of functionals: USSR Computational Mathematics and Mathematical Physics, **3**, 864–878.

Popovici, A. M., I. Sturzu, and T. J. Moser, 2015, High resolution diffraction imaging of small scale fractures in shale and carbonate reservoirs: 14th International Congress of the Brazilian Geophysical Society, H1–2–6.

Powell, M. J., 1971, Recent advances in unconstrained optimization: Mathematical Programming, **1**, 26–57.

Qian, N., 1999, On the momentum term in gradient descent learning algorithms: Neural Networks, **12**, 145–151.

Reshef, M., 2007, Velocity Analysis in the Dip-Angle Domain: Presented at the 69th EAGE Conference and Exhibition incorporating SPE EUROPEC, Expanded Abstracts.

Reshef, M., and E. Landa, 2009, Post-stack velocity analysis in the dip-angle domain using diffractions: Geophysical Prospecting, **57**, no. 5, 811–821.

Roser, M., 2013, Economic Growth: Our World in Data. (https://ourworldindata.org/economic-growth).

Sakoe, H., and S. Chiba, 1978, Dynamic programming algorithm optimization for spoken word recognition: IEEE Transactions on Acoustics, Speech, and Signal Processing, **26**, no. 1, 43–49.

Sava, P. C., B. Biondi, and J. Etgen, 2005, Wave-equation migration velocity analysis by focusing diffractions and reflections: Geophysics, **70**, U19–U27.

Schleicher, J., and J. C. Costa, 2009, Migration velocity analysis by double path-integral migration: Geophysics, **74**, no. 6, WCA225–WCA231.

Schwarz, B., 2019, Chapter one - an introduction to seismic diffraction, *in* Recent Advances in Seismology: Elsevier, volume **60** *of* Advances in Geophysics, 1–64.

Sheriff, R., and L. Geldart, 1995, Exploration seismology: Cambridge University Press.

Sherwood, J. W. C., and P. H. Poe, 1972, Continuous velocity estimation and seismic wavelet processing: Geophysics, **37**, no. 5, 769–787.

Shi, Y., X. Wu, and S. Fomel, 2020, Waveform embedding: Automatic horizon picking with unsupervised deep learning: Geophysics, **85**, WA67–WA76.

Siliqi, R., D. L. Meur, F. Gramar, L. Smith, J. P. Touré, and P. Herrmann, 2003, High-density moveout parameter fields V and η. Part one: Simultaneous automatic picking: 73rd Annual International Meeting: Society of Exploration Geophysicists Expanded Abstracts, 2088–2091.

Smyrlis, G., and V. Zisis, 2004, Local convergence of the steepest descent method in Hilbert spaces: Journal of Mathematical Analysis and Applications, **300**, 436–453.

Sripanich, Y., and S. Fomel, 2018, Fast time-to-depth conversion and interval velocity estimation in the case of weak lateral variations: Geophysics, **83**, no. 3, S227–S235.

Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, 23–48.

Symes, W. W., 1998, High frequency asymptotics, differential semblance, and velocity estimation: SEG Technical Program Expanded Abstracts 1998, 1616–1619.

Symes, W. W., 1999, All stationary points of differential semblance are asymptotic global minimizers: Layered acoustics: https://hdl.handle.net/1911/101915.

Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Taner, M. T., and F. Koehler, 1969, Velocity spectra—Digital computer derivation applications of velocity functions: Geophysics, **34**, 859–881.

Tao, Y., M. Davidson, H. Swan, S. Fomel, J. Malloy, J. Howell, S. Chiu, and R. Olson, 2012, Constrained simultaneous automatic picking for VVAZ analysis: SEG Technical Program Expanded Abstracts, 1–5.

Thomsen, L., 1988, Reflection seismology over azimuthally anisotropic media: Geophysics, **53**, no. 3, 304–313.

Thomsen, L., 2001, Seismic anisotropy: Geophysics, **66**, no. 1, 40–41.

Thomsen, L., 2002, Understanding seismic anisotropy in exploration and exploitation: SEG/EAGE Distinguished Instructor Series.

Tod, S., B. Taylor, R. Johnston, and T. Allen, 2007, Fracture prediction from wide-azimuth land seismic data in SE Algeria: The Leading Edge, **26**, no. 9, 1154–1160.

Toldi, J. L., 1989, Velocity analysis without picking: Geophysics, **54**, 191–199.

Tschannen, V., N. Ettrich, M. Delescluse, and J. Keuper, 2020, Detection of point scatterers using diffraction imaging and deep learning: Geophysical Prospecting, **68**, 830–844.

Tsvankin, I., 1997, Anisotropic parameters and P-wave velocity for orthorhombic media: Geophysics, **62**, no. 4, 1292–1309.

Tsvankin, I., 2012, Seismic Signatures and Analysis of Reflection Data in Anisotropic Media, 3rd ed.: Society of Exploration Geophysicists.

Tsvankin, I., J. Gaiser, V. Grechka, M. van der Baan, and L. Thomsen, 2010, Seismic anisotropy in exploration and reservoir characterization: An overview: Geophysics, **75**, no. 5, A15–A29.

Vail, P. R., 1977, Seismic stratigraphy and global changes of sea level.: Bull. Am. Assoc. Petrol. Geol., Mem., **26**, 49–212.

Vajnberg, M. M., 1973, Variational method and method of monotone operators in the theory of nonlinear equations: Wiley.

Ventosa, S., C. Simon, and M. Schimmel, 2012, Window length selection for optimum slowness resolution of the local-slant-stack transform: Geophysics, **77**, no. 2, V31–V40.

Vermeer, G. J. O., 2012, 3d Seismic Survey Design, 2 ed.: Society of Exploration Geophysicists.

Wu, X., and S. Fomel, 2018a, Automatic fault interpretation with optimal surface voting: Geophysics, **83**, O67–O82.

Wu, X., and S. Fomel, 2018b, Least-squares horizons with local slopes and multigrid correlations: Geophysics, **83**, IM29–IM40.

Wu, X., Z. Geng, Y. Shi, N. Pham, S. Fomel, and G. Caumon, 2020, Building realistic structure models to train convolutional neural networks for seismic structural interpretation: Geophysics, **85**, WA27–WA39.

Wu, X., and D. Hale, 2013, Extracting horizons and sequence boundaries from 3d seismic images: SEG Technical Program Expanded Abstracts 2013, 1440–1445.

Wu, X., and D. Hale, 2015, Horizon volumes with interpreted constraints: Geophysics, **80**, IM21–IM33.

Wu, Z., 1996, The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation: SIAM J. on Optimization, **6**, 748–768.

Xu, S., H. Chauris, G. Lambaré, and M. S. Noble, 2001, Common-angle migration: A strategy for imaging complex media: Geophysics, **66**, no. 6, 11877–1894.

Xue, Z., N. Alger, and S. Fomel, 2016, Full-waveform inversion using smoothing kernels: SEG Technical Program Expanded Abstracts 2016, 1358–1363.

Xue, Z., X. Wu, and S. Fomel, 2018, Predictive painting across faults: Interpretation, **6**, T449–T455.

Yan, S., and X. Wu, 2021, Seismic horizon extraction with dynamic programming: Geophysics, **86**, IM51–IM62.

Yilmaz, Ö., 2001, Seismic Data Analysis: Society of Exploration Geophysicists.

Zhang, G. J., and Z. Q. Zhang, 1998, Application of successive approximation method to the computation of the green's function in axisymmetric inhomogeneous media:

IEEE transactions on geoscience and remote sensing, **36**, 732–737.

Vita

Luke Decker was born in Seattle, Washington in 1986. After completing high school at Bishop Blanchet in 2004 he left Seattle for Walla Walla, Washington, where he enrolled at Whitman College studying physics. After graduating with his bachelor's degree in 2008 he went to work for King Canyon Buffalo, Inc. as a mudlogger working on drilling rigs across the continental United States. Early in 2011 he returned to school at Western Washington University in Bellingham, Washington. He graduated from WWU with a bachelor's degree in geophysics at the end of the year. In August 2012 he entered graduate school as a member of the Texas Consortium for Computational Seismology at the University of Texas at Austin's Bureau of Economic Geology, which he completed in the 2014. Following his masters degree he worked as a geophysicist in the Chevron Energy Technology Company until 2016 when he returned to The University of Texas at Austin for PhD studies in Computational Science, Engineering, and Mathematics at the Oden Institute for Computational Engineering and Sciences. He has completed summer internships with Anadarko Petroleum Corporation (2013), Repsol (2018), Chevron (2019), and Numerical Algorithms Group (2020), and served as the chair of the Society of Exploration Geophysicists Health, Safety, Security, and Environment committee (2017-2021). Luke has been married to Elvira Aguilera Alonso since 2016, and when he is not working or studying he loves bicycle touring.

Permanent address: Oden Institute for Computational Engineering
and Sciences
The University of Texas at Austin
201 E 24th St,
Austin, TX 78712
decker.luke@utexas.edu

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.